

A fast Bayesian approach to discrete object detection in astronomical datasets - PowellSnakes I

Pedro Carvalho,^{1*} Graça Rocha,^{2,3†} M.P. Hobson^{1‡}

¹*Astrophysics Group, Cavendish Laboratory, J.J. Thomson Avenue, Cambridge CB3 0HE, UK*

²*IPAC, Keith Spalding building, 11200 East California Blvd, Pasadena, California 91125*

³*JPL, 4800 Oak Grove Drive, Pasadena, California 91109*

Accepted - Version 7 - V—. Received —; in original form 25 July 2008

ABSTRACT

A new fast Bayesian approach is introduced for the detection of discrete objects immersed in a diffuse background. This new method, called PowellSnakes, speeds up traditional Bayesian techniques by: i) replacing the standard form of the likelihood for the parameters characterizing the discrete objects by an alternative exact form that is much quicker to evaluate; ii) using a simultaneous multiple minimization code based on Powell’s direction set algorithm to locate rapidly the local maxima in the posterior; and iii) deciding whether each located posterior peak corresponds to a real object by performing a Bayesian model selection using an approximate evidence value based on a local Gaussian approximation to the peak. The construction of this Gaussian approximation also provides the covariance matrix of the uncertainties in the derived parameter values for the object in question. This new approach provides a speed up in performance by a factor of ‘hundreds’ as compared to existing Bayesian source extraction methods that use MCMC to explore the parameter space, such as that presented by Hobson & McLachlan (2003). The method can be implemented in either real or Fourier space. In the case of objects embedded in a homogeneous random field, working in Fourier space provides a further speed up that takes advantage of the fact that the correlation matrix of the background is circulant. We illustrate the capabilities of the method by applying to some simplified toy models. Furthermore PowellSnakes has the advantage of consistently defining the threshold for acceptance/rejection based on priors which cannot be said of the frequentist methods. We present here the first implementation of this technique (Version-I). Further improvements to this implementation are currently under investigation and will be published shortly. The application of the method to realistic simulated Planck observations will be presented in a forthcoming publication.

Key words: Cosmology: observations – methods: data analysis – cosmic microwave background

1 INTRODUCTION

The detection and characterisation of discrete objects is a generic problem in many areas of astrophysics and cosmology. Indeed, one of the major challenges in such analyses is to separate a localised signal from a diffuse background. It is most common to face this problem in the analysis of two dimensional images, where one wishes to detect discrete objects in a diffuse background.

When performing this task it is often assumed that the background is smoothly varying and has a characteristic length-scale much larger than the scale of the discrete objects being sought. For example, SExtractor (Bertin & Arnouts (1996)) approximates the background emission by a low-order polynomial, which is subtracted from the image. Object detection is then performed by finding sets of connected pixels above some given threshold. Such methods run into problems, however, when the diffuse background varies on length scales and with amplitudes similar to those of the discrete objects of interest. Moreover, difficulties also arise when the rms level of instrumental noise is comparable to, or somewhat larger than, the amplitude of the localised signal one is seeking.

A specific example illustrating the above difficulties is provided by high-resolution observations of the cosmic microwave background (CMB). In addition to the CMB emission, which varies on a characteristic scale of order ~ 10 arcmin, one is often interested in detecting

* email: carvalho@mrao.cam.ac.uk

† email: graca@caltech.edu

‡ email: mph@mrao.cam.ac.uk

emission from discrete objects such as extragalactic ‘point’ (i.e. beam-shaped) sources or the Sunyaev-Zel’dovich (SZ) effect in galaxy clusters, which have characteristic scales similar to that of the primordial CMB emission. Moreover, the rms of the instrumental noise in CMB observations can be greater than the amplitude of the discrete sources. In such cases, it is not surprising that straightforward methods, such as that outlined above, often fail to detect the localised objects.

The standard approach for dealing with such difficulties is to apply a linear filter $\psi(\mathbf{x})$ to the original image $d(\mathbf{x})$ and instead analyse the resulting filtered field

$$d_f(\mathbf{x}) = \int \psi(\mathbf{x} - \mathbf{y}) d(\mathbf{y}) d^2\mathbf{y}. \quad (1)$$

Suppose one is interested in detecting objects with some given spatial template $t(\mathbf{x})$ (normalised for convenience to unit peak amplitude). If the original image contains N_{obj} objects at positions \mathbf{X}_i with amplitudes A_i , together with contributions from other astrophysical components and instrumental noise, then

$$d(\mathbf{x}) \equiv s(\mathbf{x}) + n(\mathbf{x}) = \sum_{k=1}^{N_{\text{obj}}} A_i t(\mathbf{x} - \mathbf{X}_k) + n(\mathbf{x}), \quad (2)$$

where $s(\mathbf{x})$ is the signal of interest and $n(\mathbf{x})$ is the generalised background ‘noise’, defined as all contributions to the image aside from the discrete objects. It is straightforward to design an optimal filter function $\psi(\mathbf{x})$ such that the filtered field (1) has the following properties: (i) $d_f(\mathbf{X}_k)$ is an unbiased estimator of A_i ; (ii) the variance of the filtered noise field $n_f(\mathbf{x})$ is minimised; The corresponding function $\psi(\mathbf{x})$ is the standard matched filter (see, for example, Haehnelt & Tegmark (1996)). One may consider the filtering process as ‘optimally boosting’ (in a linear sense) the signal from discrete objects, with a given spatial template, and simultaneously suppressing emission from the background.

An additional subtlety in most practical applications is that the set of objects are not all identical, but one can repeat the filtering process with filter functions optimised for different spatial templates to obtain several filtered fields, each of which will optimally boost objects with that template. In the case of the SZ effect, for example, one might assume that the functional form of the template is the same for all clusters, but that the ‘core radius’ differs from one cluster to another. The functional form of the filter function is then the same in each case, and one can repeat the filtering for a number of different scales (Herranz et al. (2002a)). Moreover, one can trivially extend the linear filtering approach to the simultaneous analysis of a set of images. Once again, the SZ effect provides a good example. Owing to the distinctive frequency dependence of the thermal SZ effect, it is better to use the maps at all the observed frequencies simultaneously when attempting to detect and characterise thermal SZ clusters hidden in the emission from other astrophysical components (Herranz et al. (2002b)).

The approaches outlined above have been shown to produce good results, but the filtering process is only optimal among the rather limited class of linear filters and is logically separated from the subsequent object detection step performed on the filtered map(s). As a result, Hobson & McLachlan ((2003); hereinafter HM03) introduced a Bayesian approach to the detection and characterisation of discrete objects in a diffuse background. As in the filtering techniques, the method assumed a parameterised form for the objects of interest, but the optimal values of these parameters, and their associated errors, were obtained in a single step by evaluating their full posterior distribution. If available, one could also place physical priors on the parameters defining an object and on the number of objects present. Although this approach represents the theoretically-optimal method for performing parametrised object detection, its implementation was performed using Monte-Carlo Markov chain (MCMC) sampling from the posterior which was extremely computationally intensive. Although considerable progress has recently been made in increasing the efficiency of sampling-based Bayesian object detection methods (Feroz & Hobson (2007)), such approaches are still substantially slower than simple linear filtering methods. Therefore, in this paper, we explore a new, fast method for performing Bayesian object detection in which sampling is replaced by multiple local maximisation of the posterior, and the evaluation of errors and Bayesian evidence values is performed by making a Gaussian approximation to the posterior at each peak. This approach yields a speed-up over sampling-based methods of many 100s, making the computational complexity of the approach comparable to that of linear filtering methods.

The outline of the paper is as follows. In section 2 we provide a brief outline of the essential Bayesian logical framework which supports the employed methodology. The previously defined entities (see section 2) are then characterized in terms of the source detection problem in section 3. A summary of the strategy to fulfil the desire goals is then suggested in section 4. In section 5 we develop the first step of the proposed algorithm: The maximisation of the posterior distribution and we give account of the errors bars on the estimated parameters of the putative source under study. A two step approach split between Fourier and real space is indicated. The second phase of the detection algorithm: The Bayesian validation procedure which we apply on the previously found likelihood maximum is explained in section 6. A prior is suggested and an upper bound on the quality of the detection is advanced. We end this section by considering a method for correcting possible systematic deviations on the evidence evaluation. In section 7 we give a detailed study on the effects that develop when adding correlation into the diffuse background that produces the substructure in which the sources are imbedded. The results from a series of simulations covering several typical scenarios are presented in section 8 and we close indicating our conclusions and possible directions for further work in section 9.

2 BAYESIAN INFERENCE

Bayesian inference methods provide a consistent approach to the estimation of a set parameters Θ in a model (or hypothesis) H for the data d . Bayes’ theorem states that

$$\Pr(\Theta|d, H) = \frac{\Pr(d|\Theta, H) \Pr(\Theta|H)}{\Pr(d|H)}, \quad (3)$$

where $\Pr(\Theta|\mathbf{d}, H) \equiv P(\Theta)$ is the posterior probability distribution of the parameters, $\Pr(\mathbf{d}|\Theta, H) \equiv L(\Theta)$ is the likelihood, $\Pr(\Theta|H) \equiv \pi(\Theta)$ is the prior, and $\Pr(\mathbf{d}|H) \equiv E$ is the Bayesian evidence.

In parameter estimation, the normalising evidence factor is usually ignored, since it is independent of the parameters Θ . This (unnormalised) posterior constitutes the complete Bayesian inference of the parameter values. Inferences are usually obtained either by taking samples from the (unnormalised) posterior using MCMC methods, or by locating its maximum (or maxima) and approximating the shape around the peak(s) by a multivariate Gaussian.

In contrast to parameter estimation problems, in model selection the evidence takes the central role and is simply the factor required to normalize the posterior over Θ :

$$\mathcal{Z} = \int L(\Theta)\pi(\Theta)d^D\Theta, \quad (4)$$

where D is the dimensionality of the parameter space. As the average of the likelihood over the prior, the evidence is larger for a model if more of its parameter space is likely and smaller for a model with large areas in its parameter space having low likelihood values, even if the likelihood function is very highly peaked. Thus, the evidence automatically implements Occam's razor: a simpler theory with compact parameter space will have a larger evidence than a more complicated one, unless the latter is significantly better at explaining the data. The question of model selection between two models H_0 and H_1 can then be decided by comparing their respective posterior probabilities given the observed data set \mathbf{d} , as follows

$$\frac{\Pr(H_1|\mathbf{d})}{\Pr(H_0|\mathbf{d})} = \frac{\Pr(\mathbf{d}|H_1)\Pr(H_1)}{\Pr(\mathbf{d}|H_0)\Pr(H_0)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0} \frac{\Pr(H_1)}{\Pr(H_0)}, \quad (5)$$

where $\Pr(H_1)/\Pr(H_0)$ is the a priori probability ratio for the two models, which can often be set to unity but occasionally requires further consideration. Unfortunately, evaluation of the multidimensional integral (4) is a challenging numerical task. Nonetheless, a fast approximate method for evidence evaluation is to model the posterior as a multivariate Gaussian centred at its peak(s) and apply the Laplace formula (see e.g. Hobson, Bridle & Lahav (2002), thereafter HBL01).

3 BAYESIAN OBJECT DETECTION

A Bayesian approach to detecting and characterizing discrete objects hidden in some background noise was first presented in an astronomical context by HM03, and our general framework follows this closely. For brevity, we will consider our data vector \mathbf{d} to denote the pixel values in a single patch in which we wish to search for discrete objects, although \mathbf{d} could equally well represent the Fourier coefficients of the image, or coefficients in some other basis.

3.1 Discrete objects in a background

Let us suppose that we are interested in detecting and characterizing some set of (two-dimensional) discrete objects, each of which is described by a template $\tau(\mathbf{x}; \mathbf{a})$. This template is defined in terms of a set of parameters \mathbf{a} that might typically denote (collectively) the position (X, Y) of the object, its amplitude A and some measure R of its spatial extent. A particular example is the circularly-symmetric Gaussian-shaped object defined by

$$\tau(\mathbf{x}; \mathbf{a}) = A \exp \left[-\frac{(x-X)^2 + (y-Y)^2}{2R^2} \right], \quad (6)$$

so that $\mathbf{a} = \{X, Y, A, R\}$. In what follows, we will consider only circularly-symmetric objects, but our general approach accommodates the templates that are, for example, elongated in one direction, at the expense of increasing the dimensionality of the parameter space.¹ In the analysis of single-frequency maps, most authors take the template $\tau(\mathbf{x}; \mathbf{a})$ to be the (pixelised) intrinsic shape of the object convolved with the beam profile. For multi-frequency data, however, it is more natural to treat the intrinsic object shape and the beam profiles separately.²

If N_{obj} objects are present in the map and the contribution of each object to the data is additive, we may write

$$\mathbf{d} = \mathbf{n} + \sum_{k=1}^{N_{\text{obj}}} s(\mathbf{a}_k), \quad (7)$$

where $s(\mathbf{a}_k)$ denotes the contribution to the data from the k th discrete object and \mathbf{n} denotes the generalised 'noise' contribution to the data from other 'background' emission and instrumental noise. Clearly, we wish to use the data \mathbf{d} to place constraints on the values of the unknown parameters N_{obj} and \mathbf{a}_k ($k = 1, \dots, N_{\text{obj}}$).

¹ Such an elongation can be caused by asymmetry of the beam, source or both. If our model assumes a circularly symmetric source but an asymmetric beam, the elongation will be constant across the map, but not if the sources are intrinsically asymmetric. In the latter case, one thus has to introduce the source position angle as a parameter, in addition to the elongation.

² In its early stages PowellSnakes treated a point source convolved with the beam as a single parameterized template. In the current version these are separate entities.

3.2 Defining the posterior distribution

As discussed in HM03, in analysing the data the Bayesian purist would attempt to infer simultaneously the full set of parameters $\Theta \equiv (N_{\text{obj}}, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{N_{\text{obj}}})$. The crucial complication inherent to this approach is that the length of the parameter vector Θ is variable, since it depends on the unknown value N_{obj} . Some sampling-based approaches are able to move between spaces of different dimensionality, and such techniques were investigated in HM03.

An alternative approach, also discussed by HM03, is simply to set $N_{\text{obj}} = 1$. In other words, the model for the data consists of just a single object and so the full parameter space under consideration is $\mathbf{a} = \{X, Y, A, R\}$, which is fixed and only 4-dimensional. Although we have fixed $N_{\text{obj}} = 1$, it is important to understand that this does *not* restrict us to detecting just a single object in the map. Indeed, by modelling the data in this way, we would expect the posterior distribution to possess numerous local maxima in the 4-dimensional parameter space, each corresponding to the location in this space of one of the objects present in the image. HM03 show this vastly simplified approach is indeed reliable when the objects of interest are spatially well-separated, and we adopt this method here.

3.3 Likelihood

In this case, if the background ‘noise’ \mathbf{n} is a statistically homogeneous Gaussian random field with covariance matrix $\mathbf{N} = \langle \mathbf{n}\mathbf{n}^T \rangle$, then the likelihood function takes the form

$$L(\mathbf{a}) = \frac{\exp \left\{ -\frac{1}{2} [\mathbf{d} - \mathbf{s}(\mathbf{a})]^T \mathbf{N}^{-1} [\mathbf{d} - \mathbf{s}(\mathbf{a})] \right\}}{(2\pi)^{N_{\text{pix}}/2} |\mathbf{N}|^{1/2}}. \quad (8)$$

Moreover, if the background is just independent pixel noise, then $\mathbf{N} = \sigma^2 \mathbf{I}$, where σ is the noise rms.

In the general case, the log-likelihood thus takes the form

$$\ln L(\mathbf{a}) = c - \frac{1}{2} [\mathbf{d} - \mathbf{s}(\mathbf{a})]^T \mathbf{N}^{-1} [\mathbf{d} - \mathbf{s}(\mathbf{a})] \quad (9)$$

where c is an unimportant constant. The first innovation in our new method is to re-cast the log-likelihood in such a way that the computational cost of evaluating it is considerably reduced. This is achieved by instead writing the log-likelihood as

$$\ln L(\mathbf{a}) = c' - \frac{1}{2} \mathbf{s}(\mathbf{a})^T \mathbf{N}^{-1} \mathbf{s}(\mathbf{a}) + \mathbf{d}^T \mathbf{N}^{-1} \mathbf{s}(\mathbf{a}), \quad (10)$$

where $c' = c - \frac{1}{2} \mathbf{d}^T \mathbf{N}^{-1} \mathbf{d}$ is again independent of the parameters \mathbf{a} . The advantage of this formulation is that the part of the log-likelihood dependent on the parameters \mathbf{a} consists only of products involving the data and the signal. Since the signal from a discrete source with parameters \mathbf{a} is only (significantly) non-zero in a limited region centred on its putative position one need only calculate the quadratic forms in (10) over a very limited number of pixels, whereas as the quadratic form in the standard version (9) must be calculated over all the pixels in the map.

In the case of independent pixel noise, the covariance matrix \mathbf{N} in (10) is diagonal, and so the quadratic forms can be calculated very rapidly. For correlated noise, however, \mathbf{N} is no longer diagonal and so evaluating the necessary elements of its inverse can be costly for a large map. Nonetheless, if the background is a statistically homogeneous random field \mathbf{N}^{-1} can be computed in Fourier space, where it is diagonal. As a result, we perform our analysis of sufficiently small patches of sky that the assumption of statistical homogeneity is reasonable. An alternative procedure would be to transform to a set of basis functions, such as wavelets, in which the data are readily compress, hence reducing the effective dimensionality of the problem; we will not, however, pursue this route.

3.4 Prior

Having determined the likelihood function, it remains only to assign a prior on the parameters $\mathbf{a} = (X, Y, A, R)$. Although not a formal requirement of our method, the simplest choice is to assume the prior is separable, so that

$$\pi(\mathbf{a}) = \pi(X)\pi(Y)\pi(A)\pi(R). \quad (11)$$

Moreover, we will assume that the priors on X and Y are uniform within the range of the image, and that priors on A and R are the uniform distributions within some ranges $[A_{\text{min}}, A_{\text{max}}]$ and $[R_{\text{min}}, R_{\text{max}}]$ respectively. This is true for our first implementation of PowellSnakes (version I). Our general approach can, in fact, accommodate more general priors. Indeed, the assumption of uniform priors on A and R is not a good one for most astrophysical problems. Typically one expects the number of discrete sources to decrease with amplitude, and their angular sizes are unlikely to be uniformly distributed. The best prior to adopt for A is one derived from existing source counts. These are typically power laws over a wide range (although care must be taken at the ends of the range to obtain a properly normalised prior). In our approach presented below, however, we will assume the log-posterior is at most quadratic in the source amplitude A , which restricts $\pi(A)$ to be of uniform, exponential or Gaussian form. (The second implementation (version II) of PowellSnakes currently under development adopts a Power law.) Turning to $\pi(R)$, existing knowledge of the angular sizes of the objects sought can be used to construct an appropriate prior, and no restriction is placed by our method on this. Of course, in the case of point sources, the overall template is simply the beam profile, the angular size of which is (usually) known in advance.³

³ Ongoing work on an improved implementation of PowellSnakes (Version II) uses the following priors: uniform priors on the position, X and Y , and scale, R , while the brightness, source amplitude A , is modelled by a Power-law: $N(s) \propto s^{-\beta}$ with β in $[2, 3]$, where s is the source flux. The range for parameter R is taken from a parameter file.

3.5 Optimal patch size. The patch border

The patches are divided into two different areas: A central area where the detection takes place and a surrounding border. The border is used to prevent the sources from being truncated by the edges of the patch since this usually impacts severely on the process of their detection and characterization. In order to avoid truncation, the border size must be at least half the largest source radial size. When enforcing this rule we are assuring that a source profile which has been cut by the patch edge is left undisturbed at least on one of its neighbour patches. When defining an optimal patch size several somewhat opposing criteria must be taken into account:

- The assumption of statistical homogeneity which always favours small patches
- Execution speed which tends to exhibit a bias towards somewhat larger patches.
 - The greater the number of patches the greater the number of border overlaps which occur. This increases the number of pixels processed more than once.
 - There is a non-negligible amount of time preparing a new patch for detection.
 - The pre-filtering stage which involves an FFT requires a power of 2 array size in order to be efficient

Our work shows that the optimal patch size is always the one which minimises the effects of the in-homogeneities of the background. However, (128 x 128) pixels patches on the NSide=1024 Healpix maps ($\sim 7.33^\circ \times 7.33^\circ$) and (256 x 256) pixels patches on NSide=2048 Healpix maps ($\sim 7.33^\circ \times 7.33^\circ$) are always good choices in terms of being a good balance between statistical homogeneity and execution speed.

4 OBJECT DETECTION STRATEGY

Once the likelihood and prior have been defined, the problem of object identification and characterization reduces simply to exploring the resulting posterior distribution

$$\ln P(\mathbf{a}) = \ln L(\mathbf{a}) + \ln \pi(\mathbf{a}), \quad (12)$$

where we have omitted an unimportant additive constant term on the right-hand side. Rather than using MCMC methods, as advocated by HM03 and Feroz & Hobson (2007), here instead we locate the local maxima of the posterior in the 4-dimensional parameter space $\mathbf{a} = \{X, Y, A, R\}$ and perform a Gaussian approximation about each peak. The former provides the estimates of the object parameters associated with that posterior peak and, as outlined below, the latter allows one to assign uncertainties to the derived parameter values. The Gaussian approximation also allows one to estimate the Bayesian evidence associated with each posterior peak. The Bayesian evidence is used to decide whether the detected peak corresponds to a real object or is just a ‘conspiracy’ of the background noise field (see Section 6).

Such an approach was also advocated by HM03, in which a simulated annealing downhill simplex minimiser was used in an iterative object detection scheme. At each iteration, the minimiser located the global maximum of the posterior and an object with the optimal parameter values was subtracted from the map before commencing the next iteration. The process was repeated until the first rejected detection.

Here we adopt a different strategy that obviates the need to attempt to locate the global maximum at each iteration, since this is very computationally expensive. Instead, our approach consists of launching a series of simple downhill minimisations. The choice of starting point for each minimisation, the minimiser employed and the number of minimisations performed are discussed in detail in Section 5. The end-point of each minimisation launched will be a local maximum in the posterior, giving the optimal parameter values for some putative detected object. A Gaussian approximation to the posterior is constructed about the peak and the detection is either accepted or rejected based on an evidence criterion (see Section 6). If the detection is accepted, then an object with the optimal parameter values is subtracted from the map before the next downhill minimisation is launched. Although our method does leave open the possibility of repeatedly detecting the same spurious sources, the speed of our algorithm assures that this is not a problem (see Section 8).

It should be noted, however, that the only reason for subtracting accepted objects from the map is to avoid multiple real detections. Moreover, when attempting to detect sources at very low signal-to-noise ratios, one might occasionally accept a spurious detection as a real source, and its subtraction from the map may damage subsequent downhill minimisations. Ideally, therefore, one should avoid subtracting sources altogether. One such approach is simply to accept that real objects may be detected repeatedly, and check for duplications in the optimal sets of parameters obtained in each accepted detection (e.g. parameters sets that are similar to well within the derived uncertainties). We have not implemented such a method for this paper, but this is under investigation.

As discussed in Section 5 determining the number of the downhill minimisations to perform is a key part of our method. However once we perform these minimizations, we ensure that no real object has been missed by performing a further set of downhill minimisations as follows. Since the spatial extent of any object in the data map must be greater than that of the beam, we divide the map into patches of area equal to the beam size. For each such patch, we then launch a downhill minimiser with initial values of X and Y equal to the coordinates of the centre of the patch, and with A and R values equal to the midpoints of their respective prior ranges. This is the case for our first implementation of PowellSnakes (version I, for white noise only). This approach is unlikely to miss any remaining peaks in the posterior associated with real objects, provided the number density of sources is sufficiently low that they are spatially well-separated. Clearly, such an approach will perform less well if two or more sources (partially) overlap, in which case a single source of large spatial extent may be fitted to them.

In the earlier stages of PowellSnakes implementation we performed the analysis completely in real space, rather than pre-filtering the map. However this procedure slowed down the execution of the algorithm. The evaluation of the objective function for each set of parameters is a very expensive operation which unfortunately lies deep in the inner loop of the code, ending up being called several hundreds of times

for each “snake”. Even a small optimization of its execution has a gigantic impact on the overall performance of the code. We initialize the Powell minimizer, (ie the tail of each of the “snakes”), with the central points of the grid defined in subsection 5.2 and the midpoints of the $[A, R]$ parameters. The first step of the minimization process is always to search the position sub-space keeping R and A constant. As we have to repeat this step each time we start a new “snake”, finding a fast way of evaluating the likelihood in all the positional sub-space available is very advantageous. This is where the filtering process steps in. Filtering with the matched filter automatically provides us with a fast evaluation of the likelihood restricted to the positional subspace. Furthermore it gives a very good estimate of the starting value for the A parameter, using the classical formula of the amplitude of a field filtered by a matched filter (16). Next, we need to find out the optimal value of R parameter which is no longer considered constant.

As a final note, one must take care when a source is located near the edge of the map, since it will be partially truncated. This is not, however, a severe limitation as one can allow an overlap of at least the large expected size when dividing the full map into small patches; a truncate source in one patch will appear intact in a neighbouring one. This strategy does result in a small extra processing overhead, since several pixels will be processed more than once, but this is a minor consideration.

5 MAXIMISING THE POSTERIOR DISTRIBUTION

In this section, we outline the method employed to locate the multiple local maxima of the posterior distribution.

5.1 Pre-filtering step

Let us begin by writing the template (6) as

$$\tau(\mathbf{x}; \mathbf{a}) = A t(\mathbf{x} - \underline{\mathbf{X}}; R), \quad (13)$$

where A , $\underline{\mathbf{X}}$ and R are, respectively, the amplitude, vector position and size of the object, and $t(\mathbf{x}; R)$ is the spatial shape of a unit height reference object of size R centred on the origin. Then the signal vector in the log-likelihood (10) is simply

$$\mathbf{s}(\mathbf{a}) = A \mathbf{t}(\underline{\mathbf{X}}, R), \quad (14)$$

where the vector \mathbf{t} has components $t_i = t(\mathbf{x}_i - \underline{\mathbf{X}}; R)$, in which \mathbf{x}_i is the position of the i th pixel in the map. Substituting the above expression into the log-likelihood, assuming a constant prior and differentiating with respect to A gives

$$\frac{\partial \ln P}{\partial A} = \mathbf{t}^T(\underline{\mathbf{X}}, R) N^{-1} [\mathbf{d} - A \mathbf{t}(\underline{\mathbf{X}}, R)], \quad (15)$$

which on setting equal to zero yields an analytic estimate for the amplitude:

$$\hat{A}(\underline{\mathbf{X}}, R) = \frac{\mathbf{t}^T(\underline{\mathbf{X}}, R) N^{-1} \mathbf{d}}{\mathbf{t}^T(\underline{\mathbf{X}}, R) N^{-1} \mathbf{t}(\underline{\mathbf{X}}, R)}. \quad (16)$$

Note that, under the assumption of statistical homogeneity, the denominator in the above expression does not depend on the source position $\underline{\mathbf{X}}$, and thus need only be calculated once (say for a source at the origin) for any given value of R . Generalising the notation of Savage & Oliver (2006), we denote this denominator by $\alpha(R)$ and the numerator in (16) by $\gamma(\underline{\mathbf{X}}, R)$. It is worth noting that the quantity $\sqrt{\alpha(R)}$ is merely a generalised signal-to-noise ratio for a unit amplitude object integrated over its spatial template.

More importantly the estimator (16) is precisely the filtered field produced by a classical matched filter, in which one assumes a value for R . Moreover, it is straightforward to show that the corresponding log-likelihood (10) can be written

$$\ln L(\underline{\mathbf{X}}, \hat{A}, R) = c' + \frac{1}{2} \alpha(R) [\hat{A}(\underline{\mathbf{X}}, R)]^2. \quad (17)$$

Thus, we show here that for a given value of R , peaks in the filtered field correspond precisely to peaks in the log-likelihood considered as a function of $\underline{\mathbf{X}}$. If the objects sought all have the same size R and this is known in advance, then the local maxima of the now 3-dimensional posterior in the space (X, Y, A) are all identified by this method. This scenario corresponds to point sources observed by an experiment with a known beam profile. However the assumption that the sizes of the objects are known in advance might lead us astray. To show that this is indeed the case, consider the following:

- Suppose two sources are so close they cannot be resolved by the optical device: The intensity profile usually still matches fairly well the of beam profile but with a much enlarged radial scale. Taking the radius of the beam profile as an unknown parameter, PowellSnakes will still identify them wrongly as a single source, but the total estimated flux will much closely match the true value of the total flux coming from both sources.
- We are assuming a statistical homogeneous background: this is never strictly true, but keeping the patches small enough this is usually a good assumption. Nevertheless, small changes in the beam profile across the patch, which in terms of the noise components (CMB, Galactic foreground, etc.) may be perfectly ignored, when estimating the sources fluxes might lead us to severely biased values.
- We are assuming point sources: this limits the range of application of this method. One of the driving motivations behind PowellSnakes is the detection of SZ clusters. This implies the capability of multi-frequency operation and a source profile with a well defined shape. This source shape template is only useful in real situations if we are to allow different spatial scales. In order to achieve this extra degree of freedom we introduced the Radius $[R]$ parameter.

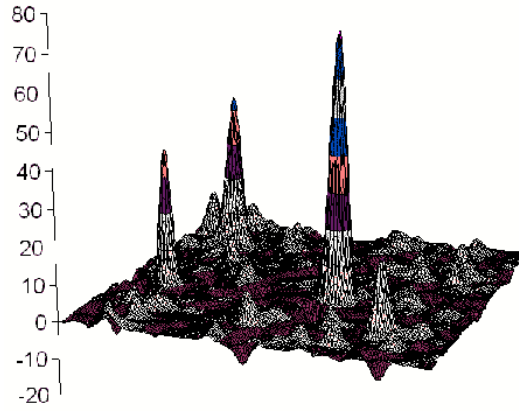


Figure 1. Log-likelihood manifold, equation (10) using template (6) (A and R dimensions suppressed); white noise background, R parameter = 4.0 pixels; a constant has been subtracted

When the source size R is not known in advance, or the sources have different sizes, most matched filter analyses (see e.g. Herranz et al. (2005)) filter the field at several predetermined values of R and combine the results in some way to obtain estimates of the parameters (X, Y, A, R) for each identified source. This approach is rather ad hoc, however, and so we will pursue a different strategy based on locating the local maxima of the posterior using multiple downhill minimisations. At first sight it might seem tempting to use the result (16) to reduce the parameter space to the 3 dimensions (X, Y, R) , since with knowledge of these parameters one can immediately calculate \hat{A} . Such an approach can, however, lead to numerical instabilities when the denominator $\alpha(R)$ in (16) is very small (which occurs in the low signal-to-noise regime) as our numerical tests confirmed. Thus we do not pursue this approach.

We do, however, make use of the result (16) by first pre-filtering the map assuming a value for R equal to the mean of its prior range. The advantage of this pre-filtering stage is that, even in the presence of objects of different sizes, the filter field often has a pronounced local maximum near many real objects in case the R parameter range is not too wide and the value of R employed in the filtering process stays close to 0 (figure 1).

5.2 2-dimensional Powell minimisation step

In the next step of our analysis, we launch $N = 100$ (see below) downhill minimisations in the two-dimensions (X, Y) using the Powell algorithm (Press et al. (1992)) with starting points chosen as described below. This method is a simple direction-set algorithm that requires only function values. The method makes repeated use of a 1-dimensional minimisation algorithm, which in our case is Brent's method; this is an interpolation scheme that alternates between parabolic steps and golden sections. In addition, we 'enhance' the basic Brent line minimizer algorithm by introducing a 'jump' procedure designed to avoid small local minima (such as those resulting in the posterior from the noise background). A description of this 'jump' procedure is given below. The reason for this 2-dimensional minimisation step is simply to obtain a good estimate of the (X, Y) -positions of the sources. As a byproduct, using formula (16), we end up obtaining an approximate estimate of the putative source amplitude A as well.

Assuming that sensitivity and experiment resolution are such that we should only expect a very small number of partially resolved sources, one may assume that in each patch of beam size area we should not find more than one real peak. The likelihood peaks in the positional sub-space (keeping A and R constant) are not expected narrower than the original beam at least when we are dealing with the white noise only case. This is so because the likelihood positional surface is nothing but the correlation of the beam shape with, on average, another beam shape eventually with a different R . Thus the total area where the correlation will have values significantly different from 0 is expected to be larger than the largest area of each one of the original beams. These assumptions naturally define the coarseness one should use when exploring the patch, namely, the beam size. As we are assuming that beam shapes with different radii may actually occur, the largest possible resolution happens when the smallest beam is considered. So far there is no prior information to eventually distinguish one part of the patch from the others, therefore we assume that a uniform rectangular grid with as many cells as the total number of "beams in the patch", would be perfectly adequate to define the starting points of the "snakes". Thus:

$$N_{\text{Snakes}} = \frac{\text{Patch total number of pixels inside the 'core' area}}{\text{Size of the smallest beam (in pixels)}} \quad (18)$$

We add an extra two dimensional random offset from the central points of each grid cell in order to avoid some unwanted correlation effects due to the perfect spacing of the grid. Some may argue that, in addition to the likelihood real peaks (those resulting from the sources in the map), we should expect a large number of fake ones resulting from the interaction of the filter with the noise fluctuations (figure 1). We considered such possibility by devising a "jump" procedure (see below) to avoid the "snakes" from being "caught" on those fake peaks. Nevertheless, as in the case of a CMB background, when the number of these spurious peaks largely outnumbers the real ones, one should increase the density of "snakes" (figure 2). For the white noise detection examples presented in section 8, we have taken all of the above into

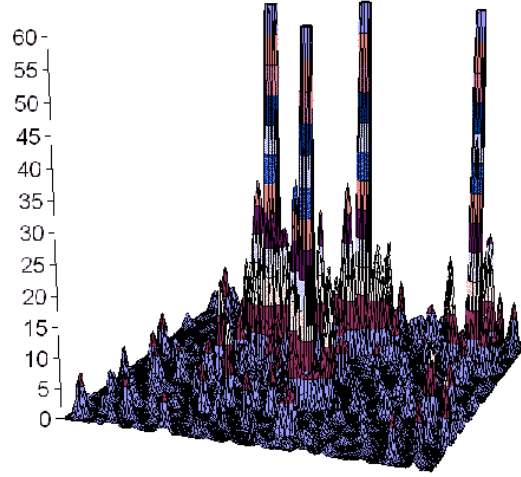


Figure 2. The likelihood ratio manifold of the models for the presence of a source over its absence $\frac{L_{H1}}{L_{H0}}$ (A and R dimensions suppressed): radius $R = 4.0$ pix; Background parameters $s_0 = 11.57$ pix, $w_I = 0.01$, $w_B = 0.99$ (see section 7)

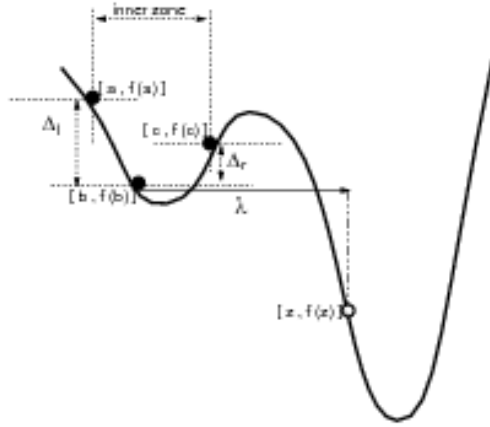


Figure 3. The jump procedure

consideration. The number of snakes chosen, $N_{\text{snakes}} = 100$, meets these requirements, for both cases (ie LISNR and HMcL, see section 8). As we assumed that no sources are placed into the patch borders we consequently did not include those pixels.

5.2.1 The jump procedure

The Brent line minimizer keeps at any time a set of three pairs of values, argument and function value, where the middle pair is said to be “bracket”. This is because the argument part of the second pair lies in the interval defined by the other two and its function value is lower than the function values of the interval limiting pairs. We call them

$$\begin{cases} [a, f(a)] & \text{Left bound} \\ [b, f(b)] & \text{Current "particle" solution} \\ [c, f(c)] & \text{Right bound} \end{cases} \quad (19)$$

Like the simulated annealing example previously referred, the rational behind the jump procedure was borrowed from natural world, but this time from the realm of quantum mechanics. We imagine our current best guess, the bracketed pair, $[b, f(b)]$, as the particle inside a potential well defined by the limiting pairs (left and right bound). If we adopt the “classical” point of view there is no way the particle could possibly escape from the bracketing barriers. This is exactly what happens with the traditional Brent procedure. Instead if we use a quantum approach it is still possible for the particle to tunnel out from the barrier.

The transmission coefficient T of a particle with energy E which faces a potential barrier with a height equal to V_0 ($V_0 > E$), and a width l , is given by:

$$T \simeq \frac{16E(V_0 - E)}{V_0^2} e^{-2\rho_2 l}, \quad (20)$$

where ρ_2 is $= \sqrt{\frac{2m(V_0 - E)}{\hbar^2}}$, m is the particle mass and \hbar is Dirac constant, where we assume $\rho_2 l \ll 1$ (Claude Cohen-Tannoudji et al. (1973)). From the quantum mechanical example we have only retained the negative exponential dependence with the barrier length (the jump distance in our jump model) and the characteristic constant depending on the difference between the particle energy and the barrier maximum potential. Following this example we define a random variable Λ exponentially distributed

$$p(\Lambda = \lambda) = \frac{1}{\theta} e^{-\lambda/\theta}, \quad (21)$$

with average θ . The next step is to define θ dependence on the parameters, bearing in mind the tunnelling effect behaviour. We chose:

$$\theta = L \left(\frac{f_s - \Delta}{f_s} \right) \quad (22)$$

where L is the “maximum average jump length”, f_s is the “function values scale”, and Δ is either the “right barrier” $\equiv \Delta_r$ when we jump forward or “left barrier” $\equiv \Delta_l$ when we are trying a jump along the opposite direction (figure 3).

The procedure has two tuning parameters: one (whose value is given as input from the parameter file) to divide the maximum allowed range for the current jump in order to produce the L value, and another one which purpose is to define a scale for the function values. This second value: f_s is the maximum range of function values evaluated from the “snake’s tails” locations (initial values drawn from the grid as explained above). When we start the minimization process by defining the bracketing interval, the well walls (defined by the differences of the function values, Δ_r and Δ_l) are usually large and consequently only very small tunnels will be allowed. If $\Delta > f_s$ no jumps will be tempted. As the minimizer converges and approaches the minimum, the barriers will become smaller and consequently bigger tunnels will become more and more probable. This jump procedure has the advantage of exploring the function domain progressively as the algorithm approaches the bottom of the well. If by accident we have first hit the global minimum when the solution is still far from the bottom only small jumps are allowed and the probability of tunnelling to a local peak is small. As soon as we get to the bottom, $\Delta \approx 0$, large tunnels become probable ($\theta \approx L$), but as the function is now close to its lowest value, a successful jump is now improbable. But if instead we have started within a local peak instead of becoming “stuck”, as the minimizer approaches its last steps, large jumps become common and the probability of “plunging” into a deeper well is now very high. One of the main reasons for choosing the exponential distribution, apart its close relationship with the quantum tunnelling, is the simplicity and efficiency of a random deviates generator whose output will follow it (provided we already have a routine which produces random deviates with a uniform probability distribution inside a certain interval). We proceed as follows:

For each Brent iteration:

- Draw a sample u from an uniform distribution between $[0, 1)$. If $u \geq 0.5$ choose a positive jump, otherwise move backwards.
- Draw from the tunnelling distribution. If the jump ends up inside the “inner zone” we reject it because we are on the same well we have started from. If the jump is outside the “inner zone” we evaluate the function at that point $[z, f(z)]$. If the new function value is lower than the one that initiated the jump, $f(z) < f(b)$, we accept it and restart the Brent minimizer with this new value.

5.2.2 The effectiveness of the jump procedure in recovering “lost snakes”

Our extensive set of simulations covering a typical assemblage of different scenarios shows that the use of the jump procedure has a fairly good effectiveness. We have achieved equal levels of quality on the obtained results using ~ 20 -25 % fewer snakes. However, within our current setup its importance is far from critical since speed is not a concern to us. The extra level of complexity which results from the introduction of the procedure, could have been avoided by launching extra “snakes”. However, when dealing with such a scenario where the cost of starting additional “snakes” is much higher or the problem cannot just be circumvented by setting up more starting points for the minimization procedure, we believe that the jump procedure might have an important role to play.

5.3 4-dimensional Powell minimisation step

Instead of just repeating the filtering procedure using different values for the R parameter, we proceed by following a rationale which tries to mimic that behind the FFT immense speed up. FFT is fast because it reduces, using factorization, the number of operations needed to compute the “complete set” of Fourier coefficients. If we only need one or two (if the number of coefficients we want to compute is less than $\ln(n)$; n being the size of the vector) then it would be more advantageous to use the traditional formula directly. Hence, we now search the full parameter space (X , Y , A and R) with the help of the Powell minimizer, but this time using the previously estimated parameters (from the 2-dimensional step) as starting values. We are not expecting the “Powell snake” to unfold significantly outside a small neighbourhood of the initial value. Thus, as we do not need to evaluate the likelihood outside this small positional range anymore (as in the FFT case where we just want to compute one or two coefficients) using the real space proves to be more advantageous.

In the next step, we perform multiple Powell minimisations in the full 4-dimensional space, starting each minimisation from the

(X, Y, A, R) - positions found in the previous step. In this step the ‘jump’ procedure described above is not used. This is one of the main reasons why we have first performed the 2-dimensional step (and the pre-filtering). From our experience we could verify that after performing the 2-dimensional step, the neighbourhood of the likelihood manifold around the vector solution $\{X, Y, A, R\}$ we had so far obtained was rather smooth and “well behaved”. Thus, the necessity for a procedure in order to avoid local maxima is no longer required. The resulting end-point of each minimisation is a local maximum of the posterior, giving the optimal parameter values for some putative detected object. A Gaussian approximation to the posterior is constructed about the peak and the detection is either accepted or rejected based on an evidence criterion (see Section 6). If the detection is accepted, then an object with the optimal parameter values is subtracted from the map before the next downhill minimisation is launched.

5.4 Error estimation - Fisher analysis

For each accepted object, estimates of the uncertainties on the derived parameters are obtained from the Gaussian approximation to the posterior around that peak. In detail the covariance matrix of the parameter uncertainties has elements $C_{ij} \equiv \langle \delta a_i \delta a_j \rangle$, which can be obtained by inverting (minus) the Hessian matrix at the peak. In the case of uniform priors, one has simply

$$(C^{-1})_{ij} = - \left\langle \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right\rangle \quad (23)$$

Here we present an estimation of the putative source parameters’ error bars using the Fisher analysis over the model H_1 . We present two different cases:

$$\begin{cases} \text{White noise} & \rightarrow \text{The analysis is carried out in real space} \\ \text{“Coloured” noise} & \rightarrow \text{Fourier space is used instead (see subsection 7.3)} \end{cases}$$

We start by defining F_{ij} as the negative expectation of the log-likelihood Hessian matrix, in other words the “Fisher information matrix”. We are assuming, as we have already stated, a Gaussian approximation of the likelihood around its maximum in parameter space.

Let $L_{H_1} \equiv \Pr(\mathbf{D}|\Theta H_1)$ be the likelihood of model H_1 , then

$$F_{ij} = - \left\langle \frac{\partial^2 \ln(L_{H_1})}{\partial \theta_i \partial \theta_j} \right\rangle_{\Theta} \quad (24)$$

Assuming a Gaussian model for the diffuse background we have:

$$-\ln(L_{H_1}) = c + \sum_{i,j} N_{ij}^{-1} d_i \tau_j - \frac{1}{2} \sum_{i,j} N_{ij}^{-1} \tau_i \tau_j - \frac{1}{2} \sum_{i,j} N_{ij}^{-1} d_i d_j, \quad (25)$$

where N_{ij}^{-1} is the (ij) element of the inverse correlation matrix of the diffuse background, p_i is the i map pixel, τ_i is the i pixel of the object template $\tau(\mathbf{r}, A, \bar{\Theta}, \mathbf{r}_0)$ and c is an unimportant constant. The parameters of the object template are: A the amplitude of the object, \mathbf{r}_0 the positional parameters and $\bar{\Theta}$ which represents the ensemble of the remaining parameters. In this analysis we are assuming a circular symmetric source. Thus, $\bar{\Theta}$ contains only a single parameter, R which provides a scale for the radial dimension of the object. ie, $R \equiv$ “Radius of the object”. We will assume, from now on, the following functional dependence of the source template on its parameters:

$$\tau(\mathbf{r}, A, \bar{\Theta}, \mathbf{r}_0) = A t \left(\frac{|\mathbf{r} - \mathbf{r}_0|^2}{R^2} \right) \quad (26)$$

5.4.1 White noise

For a background diffuse component generated from a Gaussian stationary process with coherence length much smaller than the pixel size, ie Gaussian white noise, the following condition applies:

$$N_{ij}^{-1} = \frac{1}{\sigma_i^2} \delta_{ij} \quad (27)$$

This is the situation when dealing with a system where the instrumental pixel noise is the dominant source of the background component. Adding the condition for statistical homogeneity :

$$\frac{1}{\sigma_i^2} = \frac{1}{\sigma^2}, \quad (28)$$

and substituting both conditions into the log-likelihood expression this becomes

$$-\ln(L_{H_1}) = c + \frac{1}{\sigma^2} \left(\sum_i d_i \tau_i - \frac{1}{2} \sum_i \tau_i^2 - \frac{1}{2} \sum_i d_i^2 \right), \quad (29)$$

Evaluating expression (24) for all the possible parameters' combinations and using the following sort criterion $\{A, R, X, Y\}$, one obtains

$$F_{ij} = \begin{bmatrix} \frac{\alpha \hat{R}^2}{\sigma^2} & \frac{-2\delta \hat{A} \hat{R}}{\sigma^2} & 0 & 0 \\ \frac{-2\delta \hat{A} \hat{R}}{\sigma^2} & \frac{4\beta \hat{A}^2}{\sigma^2} & 0 & 0 \\ 0 & 0 & \frac{4\gamma \hat{A}^2}{\sigma^2} & 0 \\ 0 & 0 & 0 & \frac{4\gamma \hat{A}^2}{\sigma^2} \end{bmatrix}, \quad (30)$$

where $\{\alpha, \beta, \delta, \gamma\}$ are pure numerical constants

$$\begin{cases} \alpha = 2\pi \int_0^\infty r t (r^2)^2 dr, & \beta = 2\pi \int_0^\infty r^5 t' (r^2)^2 dr \\ \gamma = \pi \int_0^\infty r^3 t' (r^2)^2 dr, & \delta = 2\pi \int_0^\infty r^3 t (r^2) t' (r^2) dr \end{cases}, \quad (31)$$

$t'(x) \equiv \frac{dt(x)}{dx}$ and $\hat{\Theta} \equiv \{\hat{A}, \hat{R}, \hat{X}, \hat{Y}\}$ are the values of the template parameters which maximize the likelihood. We have used pixel metrics together with the assumption that the sums can be evaluated using the continuous limit (this introduces only minor errors). When considering a source template with exactly the same functional dependence of equation (6), the numerical constants become:

$$\begin{cases} \alpha = \pi, & \beta = \frac{\pi}{2} \\ \gamma = \frac{\pi}{8}, & \delta = -\frac{\pi}{2} \end{cases} \quad (32)$$

Now inverting the F matrix

$$F_{ij}^{-1} = \begin{bmatrix} \frac{\beta \sigma^2}{4\hat{R}^2 \mathcal{D}} & \frac{\delta \sigma^2}{2\hat{R} \hat{A} \mathcal{D}} & 0 & 0 \\ \frac{\delta \sigma^2}{2\hat{R} \hat{A} \mathcal{D}} & \frac{\alpha \sigma^2}{4\hat{A}^2 \mathcal{D}} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{4\hat{A}^2 \gamma} & 0 \\ 0 & 0 & 0 & \frac{\sigma^2}{4\hat{A}^2 \gamma} \end{bmatrix}, \quad (33)$$

where $\mathcal{D} \equiv \alpha\beta - \delta^2$ and using the Cramèr-Rao inequality, which in this case reduces to an equality:

$$\Delta\theta_i = \sqrt{F_{\theta_i \theta_i}^{-1}}, \quad (34)$$

since we are dealing with a max-likelihood estimator which is efficient. Writing explicitly the above expression (34) for each parameter, one finally obtains the desired parameters' error bars:

$$\Delta A = \frac{\sigma}{\hat{R}} \sqrt{\frac{\beta}{4\mathcal{D}}}, \quad \Delta R = \frac{\sigma}{\hat{A}} \sqrt{\frac{\alpha}{4\mathcal{D}}}, \quad \Delta X = \frac{\sigma}{\hat{A}} \frac{1}{\sqrt{4\gamma}}, \quad \Delta Y = \frac{\sigma}{\hat{A}} \frac{1}{\sqrt{4\gamma}} \quad (35)$$

Defining $PSNR \equiv \frac{\hat{A}}{\sigma}$, "Peak Signal to Noise Ratio", and substituting into (35):

$$\Delta A = \frac{\sigma}{\hat{R}} \sqrt{\frac{\beta}{4\mathcal{D}}}, \quad \Delta R = \frac{1}{PSNR} \sqrt{\frac{\alpha}{4\mathcal{D}}}, \quad \Delta X = \frac{1}{PSNR} \frac{1}{\sqrt{4\gamma}}, \quad \Delta Y = \frac{1}{PSNR} \frac{1}{\sqrt{4\gamma}}, \quad (36)$$

we get another way of expressing the error bars, where we have emphasized their dependence on the "quality" of the signal.

6 BAYESIAN OBJECT VALIDATION

It is clear that a key component of our approach is the step for accepting/rejecting as a real object each posterior peak. Indeed, a reliable means for performing this step is crucial for the method to function properly. The decision whether to accept or reject a detection is based on the evaluation of the Bayesian evidence for two competing models for the data given by equation (5). PowellSnakes evaluates the models evidence using a Gaussian approximation to the logarithm of the likelihood ratio of the competing models

$$\ln \left(\frac{\Pr(D|\Theta H_1)}{\Pr(D|H_0)} \right), \quad (37)$$

It does so by expanding Equation (37) around its maxima in the parameter space and rejecting all the terms equal and above the third order. As the null model ($H_0 \equiv$ "There is no source centred in region S ") doesn't have any parameter, thus behaving like a constant when

considering the parameters space, maximizing the logarithm of likelihood ratio holds the same results as the max-likelihood estimator of the source parameters of the H_1 model only. Therefore finding the peaks of the logarithm of likelihood ratio, corresponds to computing the max-likelihood estimator of the putative source parameters.

The precise definition of the models H_0 and H_1 are given below. In general, however, we shall adopt a formulation of the model selection problem in which each model H_i ($i = 0, 1$) is parameterised by the same parameters $\mathbf{a} = (X, Y, A, R)$, but with different priors. In this case the evidence for each model is:

$$Z_i = \int L(\mathbf{a}) \pi_i(\mathbf{a}) d\mathbf{a}, \quad (38)$$

where the likelihood function $L(\mathbf{a})$ is the same for both models and is given by (10). Although not required by the method, we will assume that for each model the prior is separable, as in equation (11), so that for $i = 0, 1$:

$$\pi_i(\mathbf{a}) = \pi_i(\mathbf{X}) \pi_i(A) \pi_i(R). \quad (39)$$

where \mathbf{X} is the vector position of the source.

Moreover, we shall assume uniform priors within specified ranges on each parameter, although the method is easily generalised to non-uniform priors. The general form of our models for the data are:

$$H_0 = \text{'there is no source centred in the region } S',$$

$$H_1 = \text{'there is one source centred in the region } S'.$$

If the region S corresponds to the coordinate ranges $[X_{\min}, X_{\max}]$ and $[Y_{\min}, Y_{\max}]$, then

$$\pi_0(\mathbf{X}) = \pi_1(\mathbf{X}) \equiv \pi(\mathbf{X}) = \begin{cases} 1/|S| & \text{if } \mathbf{X} \in S \\ 0 & \text{otherwise} \end{cases}, \quad (40)$$

where $|S|$ is the area of the region S . We assume the prior on R is also the same for both models, so $\pi_0(R) = \pi_1(R) \equiv \pi(R)$, but the priors on A are substantially different. Guided by the forms for H_0 and H_1 given above, we take $\pi_0(A) = \delta(A)$ (which forces $A = 0$) and

$$\pi_1(A) = \begin{cases} 1/\Delta A & \text{if } A \in [A_{\min}, A_{\max}] \\ 0 & \text{otherwise} \end{cases}, \quad (41)$$

One could, of course, consider alternative definitions of these hypotheses, such as setting $H_0: A \leq A_{\lim}$ and $H_1: A > A_{\lim}$, where A_{\lim} is some (non-zero) cut-off value below which one is not interested in the identified object. We shall not, however, pursue this further here.

Given the above forms of the priors, the evidence for model H_0 is simply:

$$Z_0 = \int d\mathbf{X} dA dR \pi(\mathbf{X}) \delta(A) \pi(R) L(\mathbf{X}, A, R) \quad (42)$$

$$= L_0 \int d\mathbf{X} dR \pi(\mathbf{X}) \pi(R) = L_0, \quad (43)$$

since $L_0 \equiv L(\mathbf{X}, A = 0, R)$ is a constant and the priors are normalised. For model H_1 , the evidence reads

$$Z_1(S) = \int_S d\mathbf{X} \pi(\mathbf{X}) \int dA dR \pi(A) \pi(R) L(\mathbf{X}, A, R) \quad (44)$$

$$= \int_S d\mathbf{X} \pi(\mathbf{X}) \bar{P}_1(\mathbf{X}), \quad (45)$$

$$= \frac{1}{|S|} \int_S \bar{P}_1(\mathbf{X}) d\mathbf{X} \quad (46)$$

where we wrote explicitly the dependence of the evidence on the chosen spatial region S ; we have also defined the (unnormalised) two-dimensional marginal posterior $\bar{P}_1(\mathbf{X})$ and $|S|$ is the area of the region S .

So far we have not addressed the prior ratio $\Pr(H_1)/\Pr(H_0)$ in (5). Although this factor is often set to unity in model selection problems, one must be more careful in the current setting. For the sake of illustration and simplicity, let us assume that the objects we seek are randomly distributed in spatial position, i.e. they are not clustered. In that case, if μ_S is the (in general non-integer) expected number of objects per region of size $|S|$, then the probability of there being N objects in such a region follows a Poissonian distribution:

$$\Pr(N|\mu_S) = \frac{e^{-\mu_S} \mu_S^N}{N!}. \quad (47)$$

Thus, bearing in mind the above definitions of H_0 and H_1 , we have:

$$\frac{\Pr(H_1)}{\Pr(H_0)} = \mu_S. \quad (48)$$

Hence, the key equation (5) for model selection becomes:

$$\rho \equiv \frac{\Pr(H_1|\mathbf{d})}{\Pr(H_0|\mathbf{d})} = \mu_S \frac{Z_1(S)}{L_0} = \mu \frac{\int_S \bar{P}_1(\mathbf{X}) d\mathbf{X}}{L_0}, \quad (49)$$

where $\mu = \frac{\mu_S}{|S|}$ is the expected number sources (centres) per unit area.

There is a certain degree of freedom when choosing the level ρ must reach in order to consider that model H_1 is present (a detection) or not. This ambiguity may only be overcome by advocating decision theory (see Jaynes (2003)). From now on, we will always assume a criterion of

$$\text{symmetric loss} \equiv \text{“an undetected source is as bad as a spurious one.”} \quad (50)$$

When introducing the criterion of symmetric loss the condition for detection immediately becomes uniquely defined. One accepts the detection if

$$\rho > 1, \quad (51)$$

and rejects it otherwise. One can change our decision criteria, for instance suppose we want to put emphasis on reliability (which means we are willing to accept an increase in the number of undetected sources by a significant decrease in the fake ones), then one must change the threshold for acceptance/rejection. This means that one must change the value that ρ must reach.

The only remaining issue is the choice of the region S . Our method for locating peaks in the posterior is direct (local) maximisation, which yields the parameter values $\hat{\mathbf{a}} = (\hat{\mathbf{X}}, \hat{A}, \hat{R})$ at the peak.

In this peak-based approach, S is taken to be a region enclosing the entire posterior peak in the (\mathbf{X}) - subspace (to a good approximation). This, in fact, requires a little care. If S were taken to be the full mapped region, then the evidence $Z_1(S)$ would have contributions from many local peaks in the posterior. For each putative detection, however, we are only interested in the contribution to $Z_1(S)$ resulting from the posterior peak of interest. In practice, our Gaussian approximation to the posterior around this peak means that other peaks will not contribute to our estimate of $Z_1(S)$, thus making a virtue out of a necessity. Moreover, as we show below, provided the region S does enclose the posterior peak of interest in the (\mathbf{X}) - subspace, the resulting model selection ratio ρ will be independent of the region size $|S|$.

In the present case, using (46) and assuming uniform priors on A and R , one has:

$$Z_1(S) = \frac{1}{|S|} \int_S \bar{P}(\mathbf{X}) d\mathbf{X} = \frac{1}{|S|\Delta A \Delta R} \int L(\mathbf{a}) d^4 \mathbf{a}. \quad (52)$$

Making a 4-dimensional Gaussian approximation to the posterior about its peak $\hat{\mathbf{a}}$, one obtains

$$\rho \approx \frac{\mu_S}{L_0 |S| \Delta A \Delta R} (2\pi)^2 L(\hat{\mathbf{a}}) |\mathbf{C}(\hat{\mathbf{a}})|^{1/2}, \quad (53)$$

where the elements of (minus) the Hessian matrix \mathbf{C}^{-1} are given by (23) evaluated at the peak $\hat{\mathbf{a}}$. The above expression again ignores possible edge effects due to abrupt truncation of the posterior by the priors. In Section 6.4 we give an account of how we solved this problem. Using the expressions (10), (16) and (17) for the likelihood, one finds (in logarithms)

$$\ln \rho \approx \ln \mu_S - \ln |S| + 2 \ln(2\pi) - \ln(\Delta A \Delta R) + \frac{1}{2} \alpha (\hat{R}) [\hat{A}(\hat{\mathbf{X}}, \hat{R})]^2 + \frac{1}{2} \ln |\mathbf{C}(\hat{\mathbf{a}})|. \quad (54)$$

Most importantly, since

$$\mu_S \propto |S|, \quad (55)$$

we see that $\ln \rho$ is independent of the size $|S|$ of the region considered in the (\mathbf{X}) - subspace, provided S encloses entirely (to a good approximation) the peak of interest.

6.1 Acceptance/rejection threshold

Consider the variable ν

$$\nu \equiv \frac{\mathbf{d}^T \mathbf{M}^{-1} \mathbf{t}(\hat{\mathbf{a}})}{\sqrt{\alpha} \sigma^2}, \quad (56)$$

where $\langle n_i n_j \rangle = \sigma^2 M_{ij} = N_{ij}$ and $\sqrt{\alpha}$ is the generalised signal-to-noise ratio for a unit amplitude object (according to Savage & Oliver (2006) notation). This is the same as

$$\nu = \frac{\mathbf{d}^T \mathbf{N}^{-1} \mathbf{t}(\hat{\mathbf{a}})}{\sqrt{\mathbf{t}(\hat{\mathbf{a}})^T \mathbf{N}^{-1} \mathbf{t}(\hat{\mathbf{a}})}}, \quad (57)$$

which is the more common form one may find of $\nu \equiv$ “the normalized field amplitude” (López-Caniego et al. (2005)). “The normalized field amplitude” is the amplitude of a random stationary Gaussian field whose power spectrum $[\mathcal{B}(\eta)]$ satisfies the condition $\int_{\eta} \mathcal{B}(\eta) d\eta = 1$, ie, $\sigma = 1$. Using Parseval theorem, it is straightforward to show that the power spectrum of some background that was filtered by an un-normalized matched filter, $\mathcal{M}(\eta)$, satisfies:

$$\int_{\eta} \mathbf{M}(\eta) d\eta \simeq \mathbf{t}(\hat{\mathbf{a}})^T \mathbf{N}^{-1} \mathbf{t}(\hat{\mathbf{a}}). \quad (58)$$

Assuming that μ is neither a function of the position nor a function of the considered area, this implies that μ must be a constant. These conditions have been obtained using the principle of indifference along with the restriction (55). Integrating, one obtains:

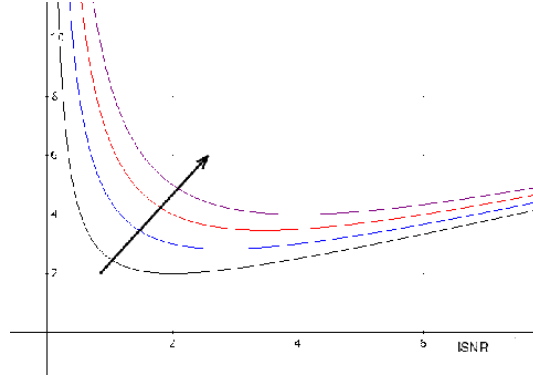


Figure 4. Right hand side of inequality (60) as function of $\text{ISNR} \equiv \hat{A}\sqrt{\alpha}$,
The arrow shows the direction of increasing \mathcal{P}

$$\mu_S = \frac{N |s|}{\Delta_s}, \quad (59)$$

where N is the expected total number of sources in the patch and Δ_s is the total area of the patch.

Let us now recover the condition for symmetric loss, $\rho > 1$, together with the key result for model selection (49). Substituting (59) into it and using expression (10) we obtain:

$$\nu > \frac{\hat{A}\sqrt{\alpha}}{2} + \frac{\mathcal{P}}{\hat{A}\sqrt{\alpha}} \quad (60)$$

where \mathcal{P} (the prior term) is:

$$\mathcal{P} \equiv \ln \left(\frac{V_T}{N 2\pi} \right) + \ln \left((2\pi)^{-1} |\mathbf{C}(\hat{\mathbf{a}})|^{-\frac{1}{2}} \right). \quad (61)$$

where V_T is the total parameter volume, ie, $V_T = (A_{max} - A_{min})(R_{max} - R_{min})\Delta_s$ and N is the expected number of sources in the map.

In Figure 4 we display the right hand side of inequality (60) as function of $\text{ISNR} \equiv \hat{A}\sqrt{\alpha}$. We plot several curves for typical values \mathcal{P} . Something which is immediately evident is that each curve has a lower bound which depends on \mathcal{P} which in turn depends on the priors. Regardless of how small the peak is there is always a lower bound for it to be considered a true detection and not merely a noise fluctuation. This is to be expected from an assessment condition which enforces a policy of robustness against false detections (spurious). The threshold curve rises on both directions: when the signal (peak) is too weak because there is a strong possibility of being a noise fluctuation; and when the signal is high because there is a lot of evidence supporting the presence of an object. Hence, rising up the threshold assures an extra level of security against spurious detections without compromising the degree of detection.

Expanding \mathcal{P} using the results from subsection 5.4 and substituting them into (61) we obtain:

$$\mathcal{P} = \ln \left(\frac{V_T \sigma^4}{32\pi^2 \sqrt{\mathcal{D}} \gamma N \hat{A}^3 \hat{R}} \right). \quad (62)$$

The curves in figure 5 show the dependency of the threshold for detection on the estimated amplitude $\hat{A} \equiv$ “estimated source amplitude”, of the putative source under study. We used data that close mimics the example from HM03: $V_T = 81000$ ($5 \times 0.5 \times 32400$), $N = 8$, $\sigma = 1$, $R \in [5, 10]$ and $A \in [0.25, 0.75]$. We plot three curves for each prior with R taking the values $\{5(\text{blue}), 7(\text{green}), 10(\text{red})\}$.

6.2 Detection and characterisation robustness

The Bayesian framework provides the necessary tools for both of the primary actions of PowellSnakes: detection and characterization. However, the impact of a mismatched prior parameter value has considerably different consequences depending on what we are doing. According to Jaynes (2003), the Bayesian inference rules prescribe that when we have a sufficiently informative likelihood the posterior is only slightly affected by the choice of the priors. Hence, one may expect that a prior mismatch will only have a minimal impact when we are dealing with bright sources. Our simulations are consistent with these predictions. But when we are tackling faint sources instead, the situation changes. Now the likelihood is no longer very informative and the priors play a part. Although, when doing parameter estimation and since we are using flat priors only, they do not affect the parameter estimates. On the other hand, something quite different happens when doing the detection step. When performing model assessment the priors must be fully normalized. Choosing a wrong prior parameter range

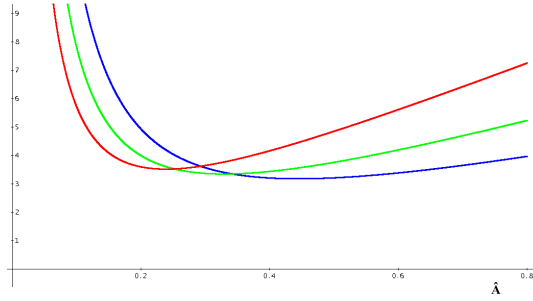


Figure 5. Right hand side of inequality (60) as function of “estimated source amplitude” $\equiv \hat{A}$. Three values of R were used (pixel units): 5 (blue), 7 (green), 10 (red).

can have an important impact on the quality of the detection, either by significantly increasing the number of false hits (spurious) or, on the contrary, making the algorithm become “blind” to detecting sources (misses).⁴

6.3 Quality of the detection: An upper bound

So far we have only discussed the condition on the detection related to the tuning of the threshold for acceptance/rejection to minimize the symmetric loss. We have said nothing about the level of success achieved. In other words, we know our procedure is tuned in order to minimize the number of undetected plus spurious sources, but we have not yet discussed how well we can do. Following Van Trees (2001) the best a “detector” can achieve is when we have absolute prior certitude about the characteristics of the detection process which is the same as knowing for sure the true values for all parameters. This condition is also known as “simple hypothesis test” and translates into the following prior:

$$\pi_0(\mathbf{a}) = \delta(a_1 - a_{10})\delta(a_2 - a_{20}) \dots \delta(a_i - a_{i0}), \quad (63)$$

where $\delta(x)$ is the Dirac delta function. Recovering expression (49), making $\mu_S = 1 \Rightarrow \Pr(H_1) = \Pr(H_0)$, which stands for, no previous information favours one model over the other \equiv “the data must say it all”, using the condition for symmetric loss (51), we have that:

$$\frac{L_1(\mathbf{a}_0)}{L_0} > 1. \quad (64)$$

Taking logarithms on both sides of the inequality and using expression (10) we obtain:

$$-\frac{1}{2}\mathbf{s}(\mathbf{a}_0)^T \mathbf{N}^{-1} \mathbf{s}(\mathbf{a}_0) + \mathbf{d}^T \mathbf{N}^{-1} \mathbf{s}(\mathbf{a}_0) > 0, \quad (65)$$

where \mathbf{a}_0 is the true parameter vector. The left hand side of the above inequality is an estimator of $\frac{L_1(\mathbf{a}_0)}{L_0}$. It is very easy to show that the likelihood ratio estimator conditioned on “there is no source”, $\frac{L_1(\mathbf{a}_0)}{L_0} |_{H_0} \equiv \Gamma_{|H_0}$ is Gaussian distributed with $N(-\frac{1}{2}ISNR_0^2, ISNR_0^2)$, where $ISNR_0 \equiv A_0\sqrt{\alpha_0}$. Hence

$$\nu_{|H_0} \equiv \frac{\Gamma_{|H_0} + \frac{1}{2}ISNR_0^2}{ISNR_0} \sim N(0, 1). \quad (66)$$

Substituting $\nu_{|H_0}$ into inequality (65), we obtain

$$\nu_{|H_0} > \frac{1}{2}ISNR_0. \quad (67)$$

Each time this inequality is satisfied we presume we are in the presence of a source, a true detection. But in our assumptions we have assumed the opposite. Hence, this is the condition for a spurious detection. Therefore, the probability for the occurrence of a spurious detection is:

$$\Pr(\widehat{H_1} | H_0) = \frac{1 - \text{erf}\left(\frac{\sqrt{2} ISNR_0}{4}\right)}{2}, \quad (68)$$

where $\widehat{H_i}$ means “once we have chosen H_i ” and $\text{erf}(x)$ is the “Gauss error function”. Since the loss is symmetric, the probability for the other type of error $\Pr(\widehat{H_0} | H_1)$, the probability that an undetected source occurs, should match that of the spurious one. Thus,

⁴ In PowellSnakes (Version II) the flat prior on the source amplitude A has been replaced by a much more realistic power law: $N(s) \propto s^{-\beta}$, and the extreme sensitivity of the detection step upon the values of the prior parameters was gone

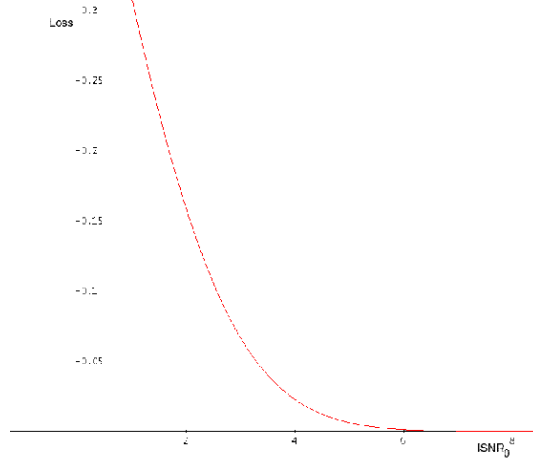


Figure 6. Curve for the theoretical lower bound for $\Pr(\text{Err}) \equiv \text{Loss}$ as a function of $ISNR_0$

$$\Pr(\widehat{H_0} | H_1) = \Pr(\widehat{H_1} | H_0) = \frac{1 - \text{erf}\left(\frac{\sqrt{2} ISNR_0}{4}\right)}{2}. \quad (69)$$

We are now in position to establish the lower bound for a detection. When performing the detection two independent and exclusive types of errors can occur:

$$\begin{aligned} \text{Spurious} &\Rightarrow \Pr(\widehat{H_1} | H_0) = \Pr(\widehat{H_1} | H_0) \Pr(H_0) \\ \text{Undetected} &\Rightarrow \Pr(\widehat{H_0} | H_1) = \Pr(\widehat{H_0} | H_1) \Pr(H_1). \end{aligned} \quad (70)$$

The total probability for an error to occur is

$$\Pr(\text{Err} | \mathbf{a}_0) = \Pr(\widehat{H_1} | H_0) \Pr(H_0) + \Pr(\widehat{H_0} | H_1) \Pr(H_1), \quad (71)$$

where the conditioning in \mathbf{a}_0 explicitly expresses the assumption of perfect prior knowledge on the parameters values (“*simple hypothesis*”).

Taking advantage of the symmetry $\Pr(\widehat{H_0} | H_1) = \Pr(\widehat{H_1} | H_0)$ and the normalization condition $\Pr(H_0) + \Pr(H_1) = 1$, we finally get

$$\text{Loss}_0 \equiv \Pr(\text{Err} | \mathbf{a}_0) = \frac{1 - \text{erf}\left(\frac{\sqrt{2} ISNR_0}{4}\right)}{2}. \quad (72)$$

Figure 6 shows the dependence of the loss theoretical lower bound on $ISNR_0$. One can see that $ISNR_0$ plays a pivotal role in defining an upper level of the quality of the detection process.

In many situations the radius of the source intensity profile, as recorded in the pixels map, may be considered known and constant throughout the patch. One of these cases is that of considering point shapeless sources. In this case one can further assume, with a high degree of accuracy, that the pixelized intensity closely follows the PSF of the antenna. Thus, considering a statistically homogeneous background, the $ISNR_0$ becomes a function of the source intensity only, A_0 , as $\alpha_0 = \mathbf{t}(R_0)^T \mathbf{N}^{-1} \mathbf{t}(R_0)$ may now be considered constant. This separation allows us to make predictions about the minimum flux that one can reliably detect taking into consideration our goals and the experimental setup. Let us define “*Normalized Integrated Signal to Noise Ratio*” $\equiv NISNR$ which is the $ISNR$ of a unit amplitude source template.

$$NISNR \equiv \sqrt{\alpha} = \sqrt{\mathbf{t}(R)^T \mathbf{N}^{-1} \mathbf{t}(R)} \quad (73)$$

This quantity is of considerable importance when establishing a boundary on the source fluxes that can be reliably detected, as it provides a scale of measure. One may think of A (the source amplitude) as being the $ISNR$ measured in $NISNR$. For a background correlation matrix considered diagonal (white noise, eq 27) and the Gaussian source template (6), an analytical evaluation of the $NISNR_0$ is possible:

$$NISNR_0 = \frac{\sqrt{\pi} R_0}{\sigma}. \quad (74)$$

In all interesting scenarios one rarely has an allowed parameter domain which sums up to single point in parameter space (“*simple hypothesis*”). Now the question is how to define a single number which somehow describes a representative figure for the detection quality

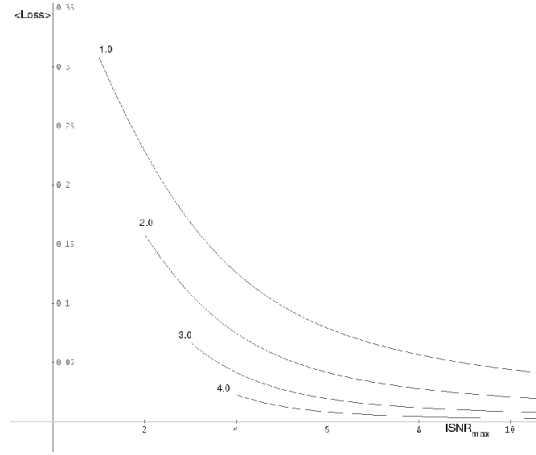


Figure 7. Plot of equation (77). Each curve represents a different value of I_m shown close to it

upper bound of the ensemble of “*simple hypothesis*” that constitute the volume of the parameter domain. Within a Bayesian framework we tackle this issue by treating the simple hypothesis parameters as nuisance parameters and integrating them out. Hence, our proposed solution is to average the “*simple hypothesis*” loss limit, $Loss_0$, over the properly normalized probability priors for the parameters:

$$\langle Loss \rangle \equiv \Pr(Err) = \int_{\mathbf{v}} \Pr(Err|\mathbf{a}) \pi(\mathbf{a}) d^N \mathbf{a}, \quad (75)$$

Substituting (72) into the equation (75) above, and dropping the zero subscript once it no longer carries any special meaning, we obtain:

$$\langle Loss \rangle = \int_{\mathbf{b}} \frac{1 - \operatorname{erf}\left(\frac{\sqrt{2} \operatorname{ISNR}(\mathbf{b})}{4}\right)}{2} \pi(\mathbf{b}) d^N \mathbf{b}, \quad (76)$$

where \mathbf{b} stands for the entire parameter set except the position coordinates. We are implicitly assuming a factorisable parameter prior (see formula 11). The above formula plays a key role when defining the expected quality for a given detection and measurement setup. If the predicted detection scenario has a low average ISNR , we should never expect a good detection performance. By experience we have verified that when tackling real situations the expected loss is usually much larger than the predicted theoretical bound.

When considering the case where the prior on the source amplitude, $\pi(A)$, can be assumed flat on the region of detection, and the source radius constant across the patch, to a good approximation an analytical evaluation of the above expression is possible:

$$\langle Loss \rangle = \frac{1}{2(I_M - I_m)} \left[u \left(1 - \operatorname{erf}\left(\frac{\sqrt{2} u}{4}\right) \right) - \sqrt{\frac{2}{\pi}} e^{-u^2/8} \right]_{u=I_m}^{u=I_M}, \quad (77)$$

where $I_m \equiv$ “minimum ISNR” and $I_M \equiv$ “maximum ISNR”. From figure 7 one realises the importance of previewing the ISNR limits of our measurement setup in order to define the expected upper bound on the detection quality.

6.4 Discussion on truncation of the posterior by the prior

So far the procedure described (formula 54) does not take into account the fact that the prior might abruptly truncate the posterior before it falls (largely) to zero, and hence may lead to an overestimation of the integral. The purpose of this discussion is essentially directed towards the truncation of the posterior which happens in the $\{A, R\}$ sub-space. When considering the position sub-space, the prior is independent of the area considered $|s|$, as far as this fully encompasses the likelihood peak, which is always the case when dealing with real scenarios.

When we were dealing with the source shape and the beam melted into a single template of radius R embedded into a background of white noise only, the effect of truncation was already evident, especially when were tackling cases of low and very low ISNR ($\equiv \hat{A}\sqrt{\alpha}$), but not extremely severe. This has provided the rational behind the “*tuning parameter*” see below. The effect was not severe because the extent of the area under the likelihood fitted fairly well inside the prior box because the A and R parameters error bars were comparatively small.

6.4.1 The tuning parameter ϕ

When assuming flat priors for all the parameters, the priors become constants and therefore can be taken out of the evidence integrals. Thus, one may now use the approximate formula (54) as a fast way of evaluating the evidence integrals, as far as the truncation effects are kept

Table 1. Tuning parameter ϕ as function of $\langle \text{ISNR}_0 \rangle$

$\langle \text{ISNR}_0 \rangle$ is small (≤ 5)	$\phi \in (0.66, 1]$
$5 < \langle \text{ISNR}_0 \rangle < 7$	$\phi \approx 1$
$\langle \text{ISNR}_0 \rangle$ is large (≥ 7)	$\phi \in (1, 1.33]$

under control. But especially when dealing with faint sources we should expect these effects to become significant and hence they must be taken into consideration. Therefore, we added a tuning parameter, ϕ , to correct for possible systematic deviations from the true value of the integral, ie:

$$\frac{\phi (2\pi)^{M/2} |\mathbf{C}(\hat{\mathbf{a}})|^{1/2}}{V_T} L(\hat{\mathbf{a}}) = \frac{\int_{\mathbf{a}} L(\mathbf{a}) d^M \mathbf{a}}{V_T} \quad (78)$$

The new expression for the assessment level curve (right hand side of inequality 60) which results from the introduction of the “*tuning parameter*”, is just the old form where we have only changed the \mathcal{P} expression:

$$\mathcal{P} = \ln \left(\frac{V_T \sigma^4}{\phi 32\pi^2 \sqrt{\mathcal{D}} \gamma N \hat{A}^3 \hat{R}} \right). \quad (79)$$

6.4.2 How to evaluate the parameter ϕ

In the simple case of white noise it is possible to compute an average value of the tuning parameter:

$$\left\langle \frac{\int_{V_{\mathbf{a}}} L(\mathbf{a}) d^M \mathbf{a}}{L_0} \right\rangle_{H_1} = \phi (2\pi)^{M/2} |\mathbf{C}(\hat{\boldsymbol{\theta}})|^{1/2} \left\langle \frac{L(\hat{\mathbf{a}})}{L_0} \right\rangle_{H_1}, \quad (80)$$

where, $V_{\mathbf{a}} \equiv$ “*Parameter volume of a single peak*”. Substituting expression (10) into (80) we get

$$\phi = \frac{\int_{V_{\mathbf{a}}} \exp \left(\frac{\mathbf{s}_0^T \mathbf{M}^{-1} \mathbf{s}(\mathbf{a})}{\sigma^2} - \frac{1}{2} \frac{\mathbf{s}(\mathbf{a})^T \mathbf{M}^{-1} \mathbf{s}(\mathbf{a})}{\sigma^2} \right) d^4 \mathbf{a}}{(2\pi)^2 |\mathbf{C}(\hat{\mathbf{a}})|^{1/2} \exp(\frac{1}{2} \hat{\alpha})}, \quad (81)$$

where “Average value of the data under hypothesis $H_1 \equiv \langle \mathbf{d} \rangle_{|H_1}$ ” = “signal vector true value $\equiv \mathbf{s}_0$ ”. The integral in the numerator is difficult to compute, but restricting our analysis to assuming a white noise background only $\mathbf{M}^{-1} = I$, obviates a numerical evaluation of this integral.

$$\phi = \frac{\int_{V_{\mathbf{a}}} \exp \left(\frac{\mathbf{s}_0^T \mathbf{s}(\mathbf{a})}{\sigma^2} - \frac{1}{2} \frac{\mathbf{s}(\mathbf{a})^T \mathbf{s}(\mathbf{a})}{\sigma^2} \right) d^4 \mathbf{a}}{(2\pi)^2 |\mathbf{C}(\hat{\mathbf{a}})|^{1/2} \exp(\frac{1}{2} \hat{\alpha})}. \quad (82)$$

From expression (82), one sees that the value of ϕ depends on \mathbf{s}_0 , the true shape of the object. In Table 1 we display several values of ϕ which result from the application of the above formula to typical scenarios ($\text{ISNR}_0 \equiv A_0 \sqrt{\alpha_0}$). The values we obtained for the “tuning parameter” are consistent with those we first expected. When dealing with scenarios with faint signals \Rightarrow *low* $\langle \text{ISNR}_0 \rangle$ we anticipate the parameter values being close to the priors inferior bounds and at the same time large error bars are expected, hence the likelihood truncation effects should make a significant contribution to the value of the evidence integral. Thus, the low value of ϕ should not be a surprise. When the opposite happens \Rightarrow *high* $\langle \text{ISNR}_0 \rangle$ the Gaussian approximation to the likelihood peak is no longer accurate and the large ϕ is just a consequence of the wider tails of true likelihood function.

6.4.3 Impact of ϕ on the performance of the algorithm

Using a series of simulations (see section 8) for different detection scenarios, we estimated the impact of the tuning parameter, ϕ , on the performance of the algorithm. We concluded that:

- The algorithm is not very sensitive to changes of ϕ .
- The differences in performance never exceeded 1.5% when compared with the fiducial value $\phi = 1$.
- The largest difference occurred when we dealt with very small values of $\langle \text{ISNR}_0 \rangle$ (< 4), or very large (> 8).
- And most important of all: “*The exact values of ϕ always produced the best results*”.

Therefore the golden rule displayed in Table 2 should be applied.

Table 2. Tuning parameter ϕ ; The golden rule

$\langle \text{ISNR}_0 \rangle \leq 5$	use $\phi = 0.75$
$5 < \langle \text{ISNR}_0 \rangle < 7$	use $\phi = 1.00$
$\langle \text{ISNR}_0 \rangle \geq 7$	use $\phi = 1.15$

7 INTRODUCING “COLOUR”

When dealing with real astronomical detection problems, the background diffuse component of the maps is not usually dominated by the instrumental noise. Nowadays, the most common scenario is the one where the largest signal component of the background has an astronomical origin: The ubiquitous CMB and the galactic foregrounds (free-free, dust, synchrotron). Currently we are focusing on the CMB especially because it can be considered statistically homogeneous and isotropic across the entire sphere with great accuracy (assumptions of our simplified model). The CMB radiation field has already an intrinsic coherence length ($\sim 10'$), which taking into consideration the current surveys resolution (usually lower than $1'$ per pixel), may be hardly considered white. In addition, as it is collected by an antenna with a finite aperture (usually several pixels), whose net effect, in Fourier space, is that of a low pass filter, an extra degree of correlation is injected into the background. Hence, only an algorithm which appropriately deals with a background whose power spectrum is not flat could possibly cope with such a setup. The power spectrum of the simplified background model we proposed (see 86) is not yet fully realistic. However, despite of its simplicity, this model, already introduces the main characteristics of this background. This allow us a direct comparison between the model predictions and the simulations results. When dealing with a background assumed to be generated by a stationary Gaussian process but now with an arbitrary coherence length, the condition (27) on the inverse correlation matrix no longer holds. Nevertheless, despite the fact the correlation matrix is no longer diagonal, if the condition of statistical homogeneity still holds, one may always assume the background correlation matrix to be circulant. Taking advantage of the circulant property of the correlation matrix and transforming to Fourier space, the problem significantly simplifies as the power spectrum of a circulant correlation matrix is always diagonal. This is not a severe limitation because considering patches of small size only, the condition for statistical homogeneity is a good approximation. We are assuming not only statistical homogeneity but isotropy as well. Most of the formulas presented here will hold even if we drop the condition for statistical isotropy. We shall ignore this fact and we will use the statistical isotropy condition throughout the remaining text. We do so because it gives us the opportunity for significant simplifications without restricting the opportunity for real applications, as this condition holds pretty well when tackling the great majority of the problems of interest. Expressing equation (25) (multiplied by -1) in Fourier space one obtains

$$\ln(L_{H_1}) = c' + \int_{\eta} \frac{\tilde{d}(\eta)\tilde{\tau}(-\eta)}{\mathcal{B}(\eta)} d\eta - \frac{1}{2} \left(\int_{\eta} \frac{\tilde{\tau}(\eta)\tilde{\tau}(-\eta)}{\mathcal{B}(\eta)} d\eta + \int_{\eta} \frac{\tilde{d}(\eta)\tilde{d}(-\eta)}{\mathcal{B}(\eta)} d\eta \right), \quad (83)$$

where the usual $\mathbf{k} = 2\pi\eta$. In a similar way for the likelihood L_{H_0} of the background

$$\ln(L_{H_0}) = c' - \frac{1}{2} \left(\int_{\eta} \frac{\tilde{d}(\eta)\tilde{d}(-\eta)}{\mathcal{B}(\eta)} d\eta \right), \quad (84)$$

where each of the symbols $\tilde{\tau}$, \tilde{d} are the Fourier transforms of τ and d respectively, and $\mathcal{B}(\eta)$ is the normalized Power Spectrum of the background, ie, $\int_{\eta} \mathcal{B}(\eta) d\eta = 1 \Leftrightarrow \sigma = 1$. In what follows the symbols with a tilde on top mean the Fourier transform of the original symbol (without the tilde). We are assuming that a pre-normalization of the map's pixels has already been done by dividing the value of each map's pixels by the map rms value ($\equiv \sigma$), which implies a value of $\sigma = 1$ for the resulting map.

Let us now introduce, for the first time, correlation effects in the background diffuse component. This correlation is to be expected not only because it already exists in the original astronomical background (CMB), and in the diffuse foregrounds (galactic components), but also due to the effect of the antenna PSF. We are considering now a simplified model for the autocorrelation function of the background, after being passed through the antenna. This model is of the form:

$$\langle \mathbf{x}_j \mathbf{x}_i \rangle_b \equiv \mathbf{n}_b(|\mathbf{x}_j - \mathbf{x}_i|) = e^{-\frac{1}{2} \frac{(\mathbf{x}_j - \mathbf{x}_i)^2}{s_0^2}}, \quad (85)$$

where s_0 is a measure of the “coherence” length of the Gaussian isotropic homogeneous stationary background field (see Goodman (2000))⁵. The autocorrelation function is already properly normalized as required by our previous assumptions, ie, $\sigma_b^2 = \langle \mathbf{x}_i \mathbf{x}_i \rangle_b = \mathbf{n}_b(0) = 1$. Transforming to Fourier space and using the Wiener-Khinchin theorem (Goodman (2000)) we get:

$$\mathcal{B}_b(\eta) = \frac{1}{2\pi\eta_0^2} e^{-\frac{1}{2} \left(\frac{\eta}{\eta_0}\right)^2}, \quad (86)$$

where $\mathcal{B}(\eta)$ is the “Power spectrum” and $\eta_0 = (2\pi s_0)^{-1}$.

⁵ There is not a complete consensus about the exact definition of “coherence length”. We shall follow Goodman’s definition, “coherence length” $\equiv \sqrt{\text{coherence area}}$, which applied to our particular form of the autocorrelation function gives $\sqrt{\pi} s_0$

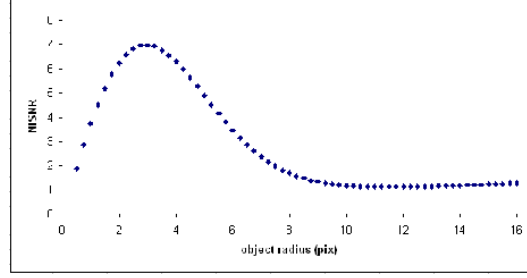


Figure 8. NISNR (σ units) vs “source radius” R (pix) ; $s_0 = 11.57$ pix, $w_I = 1/5$, $w_B = 4/5$

Together with the astronomical components one should always expect the presence of “*pixel noise*”. In our signal model, “*pixel noise*” stands for all the noise components resulting from the instrumental setup. We are assuming the “*pixel noise*” to be independent from pixel to pixel, hence configuring a white noise process. We also assume that the “*pixel noise*” is uncorrelated with all the astronomical components, thus allowing us to write the total background power spectrum as the sum of these two components:

$$\mathcal{B}(\eta) = w_B \frac{1}{2\pi\eta_0^2} e^{-\frac{1}{2}\left(\frac{\eta}{\eta_0}\right)^2} + w_I \quad (87)$$

where $\mathcal{B}_I(\eta) = 1$ (white noise process) and we added the constants $\{w_I, w_B\}$ in order to allow the signal components to be present in different amounts, hence encompassing a broader range of possible scenarios. These constants must satisfy the normalization condition $w_I + w_B = 1$ as the total power spectrum must integrate to unity.

7.1 ISNR - the symmetric loss lower bound

Let us begin by expressing in Fourier space, the equivalent representation of the ISNR, using the appropriate form of the Parseval theorem:

$$ISNR(\mathbf{a})^2 = \tau(\mathbf{a})\mathbf{N}^{-1}\tau(\mathbf{a}) = \int_{\eta} \frac{\tilde{\tau}(\eta, \mathbf{a})\tilde{\tau}(-\eta, \mathbf{a})}{\mathcal{B}(\eta)} d\eta. \quad (88)$$

First we will handle the simplest possible case where the radii of the sources are constant. We start by evaluating the $NISNR$:

$$NISNR(\mathbf{a}) = \sqrt{\int_{\eta} \frac{\tilde{t}(\eta, \mathbf{a})\tilde{t}(-\eta, \mathbf{a})}{\mathcal{B}(\eta)} d\eta} \quad (89)$$

as this will provide us with a metric to predict how deep we can go in the flux scale without compromising the detection quality. Substituting the expression (87) for the power spectrum and the unity amplitude Gaussian template (6) into (89) and evaluating the integral, (which has numerical solution only) and using typical values for the sources and background parameters we obtain the function plotted in figure 8.

When we were dealing with a white noise only background the $NISNR$ was proportional to the radii of the sources (see formula 74). However when we introduce the correlation, a completely different curve results. Now, the dependence of the $NISNR$ on the object radius stops being linear and a complex shape emerges.

7.2 The “whitening” equalizer

A quantity which plays a central role in the detection process is the likelihood ratio of the competing models (see formulas 83 and 84). After taking logarithms one obtains:

$$\ln \left(\frac{L_{H_1}}{L_{H_0}} \right) = \int_{\eta} \frac{\tilde{d}(\eta)\tilde{\tau}(-\eta)}{\mathcal{B}(\eta)} d\eta - \frac{1}{2} \int_{\eta} \frac{\tilde{\tau}(\eta)\tilde{\tau}(-\eta)}{\mathcal{B}(\eta)} d\eta \quad (90)$$

Now let us assume that our source template may be modelled by the following expression which is a generalization of equation (6):

$$\tilde{\tau}(\eta) = \tilde{h}(b_{\beta}, \eta) e^{-i2\pi\eta \cdot \mathbf{r}_0}, \quad (91)$$

where $b_{\beta} \equiv$ all variables but position, $i = \sqrt{-1}$, $\mathbf{r}_0 \equiv (x_0, y_0)$ and $\tilde{h}(b_{\beta}, \eta)$ is the Fourier transform of the source template centred in the origin. Substituting into (90) we get

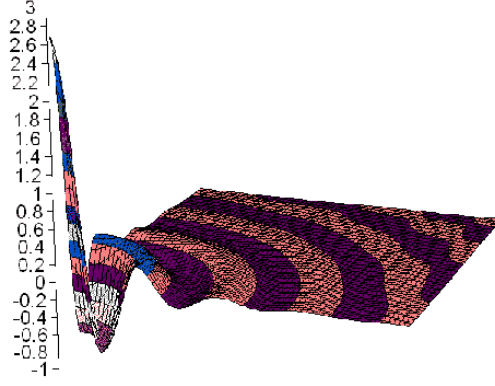


Figure 9. “whitened source” $R \ll s_0$: radius $R = 3.4$, $NISNR \approx 23.31$; Background parameters $s_0 = 11.57$, $w_I = 0.01$, $w_B = 0.99$; all distances in pixels

$$\ln \left(\frac{L_{H_1}}{L_{H_0}} \right) = \mathcal{F}^{-1} \left\{ \frac{\tilde{d}(\boldsymbol{\eta}) \tilde{h}(b_\beta, \boldsymbol{\eta})}{\mathcal{B}(\boldsymbol{\eta})} \right\}_{r_0} - \frac{1}{2} \int_{\boldsymbol{\eta}} \frac{\tilde{h}(b_\beta, \boldsymbol{\eta})^2}{\mathcal{B}(\boldsymbol{\eta})} d\boldsymbol{\eta}, \quad (92)$$

where \mathcal{F}^{-1} is the inverse bi-dimensional Fourier transform. One should already be familiar with this expression: the second term is $1/2$ of the $ISNR$ expressed in Fourier space, which doesn’t depend on the position coordinates, and the first, in the argument of the inverse Fourier transform, is the “*Linear matched filter*” applied to the map, where:

$$\text{“Linear matched filter”} \equiv \frac{\tilde{h}(b_\beta, \boldsymbol{\eta})}{\alpha \mathcal{B}(\boldsymbol{\eta})}. \quad (93)$$

Further we can read this expression in a somewhat different manner. Let us write the power spectrum as the product of its square root (the expected amplitude spectrum), ie $\mathcal{B}(\boldsymbol{\eta}) = \sqrt{\mathcal{B}(\boldsymbol{\eta})} \sqrt{\mathcal{B}(\boldsymbol{\eta})}$. Substituting into the argument of the Fourier transform (first term of 92) we have

$$\frac{\tilde{d}(\boldsymbol{\eta})}{\sqrt{\mathcal{B}(\boldsymbol{\eta})}} \frac{\tilde{h}(b_\beta, \boldsymbol{\eta})}{\sqrt{\mathcal{B}(\boldsymbol{\eta})}}. \quad (94)$$

Evaluating the square of the first term of (94) ensemble average, ie its power spectrum we obtain:

$$\left\langle \frac{\tilde{d}(\boldsymbol{\eta})^2}{\mathcal{B}(\boldsymbol{\eta})} \right\rangle = 1. \quad (95)$$

Hence, after applying the linear filter $\mathcal{B}(\boldsymbol{\eta})^{-1/2}$ we removed the “*colour*” from the background transforming it into a “*white*” Gaussian random field. This is the reason why one calls the linear matched filter a “*whitening equalizer*”. Let us name the resulting map after being “*whitened*”:

$$\text{“whitened map”} \equiv \tilde{\omega}(\boldsymbol{\eta}) \equiv \frac{\tilde{d}(\boldsymbol{\eta})}{\sqrt{\mathcal{B}(\boldsymbol{\eta})}}. \quad (96)$$

Of course, the sources buried in the map become “*whitened*” as well. In fact the second term of (94) is nothing else but a whitened source:

$$\text{“whitened source”} \equiv \tilde{\psi}(b_\beta, \boldsymbol{\eta}) \equiv \frac{\tilde{h}(b_\beta, \boldsymbol{\eta})}{\sqrt{\mathcal{B}(\boldsymbol{\eta})}}. \quad (97)$$

Rewriting formula (92) using these new definitions

$$\ln \left(\frac{L_{H_1}}{L_{H_0}} \right) = \mathcal{F}^{-1} \left\{ \tilde{\omega}(\boldsymbol{\eta}) \tilde{\psi}(b_\beta, \boldsymbol{\eta}) \right\}_{r_0} - \frac{1}{2} \int_{\boldsymbol{\eta}} \tilde{\psi}(b_\beta, \boldsymbol{\eta})^2 d\boldsymbol{\eta}. \quad (98)$$

It can be easily verified that this formula, where we used the “*whitened*” counterparts of our entities, reduces to the familiar white-noise likelihood ratio (formulas 64 and 65) expressed for convenience in Fourier space. Hence, everything we said for the white noise case applies here, even when we ought to consider the effects of correlation in the background, as long as we use the “*whitened*” entities instead.

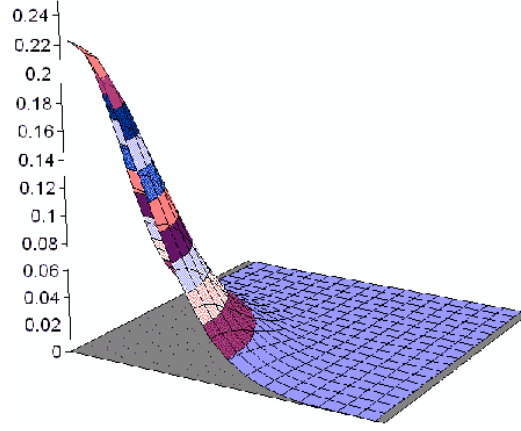


Figure 10. “whitened source” $R \approx s_0$: radius $R = 3.4$, $NISNR \approx 1.47$; Background parameters $s_0 = 2.0$, $w_I = 1/5$, $w_B = 4/5$; all distances in pixels

7.3 Fisher Analysis in Fourier space

Proceeding with the same line of reasoning as that initiated in subsection 5.4, but this time expressing L_{H_1} using Fourier space (83), we start by splitting the evaluation of the Fisher matrix into three different cases:

- The first case refers to those coefficients whose differentiation variables b_i are not positional. We shall collectively name them as $\equiv C$.

$$\frac{\partial^2 \{-\ln(L_{H_1})\}}{\partial b_i \partial b_j} = - \int_{\boldsymbol{\eta}} \mathcal{B}(\boldsymbol{\eta})^{-1} \frac{\partial \tilde{h}(b_\beta, \boldsymbol{\eta})}{\partial b_i} \cdot \frac{\partial \tilde{h}(b_\beta, \boldsymbol{\eta})}{\partial b_j} d\boldsymbol{\eta}. \quad (99)$$

These elements of the F matrix will be generally $\neq 0$.

- The second case deals with those coefficients where one of the differentiation variables belongs to set b_i and the other to x_i , the position set.

$$\frac{\partial^2 \{-\ln(L_{H_1})\}}{\partial b_i \partial x_k} = i2\pi \int_{\boldsymbol{\eta}} \eta_{x_k} \mathcal{B}(\boldsymbol{\eta})^{-1} \frac{\partial \tilde{h}(b_\beta, -\boldsymbol{\eta})}{\partial b_i} \cdot \tilde{h}(b_\beta, \boldsymbol{\eta}) d\boldsymbol{\eta}. \quad (100)$$

One may rewrite the expression (100) taking advantage of the symmetries of the expressions involved:

$$= -4\pi \int_{\eta > 0} \eta_{x_k} \mathcal{B}(\boldsymbol{\eta})^{-1} \text{Im} \left[\frac{\partial \tilde{h}^*(b_\beta, \boldsymbol{\eta})}{\partial b_i} \cdot \tilde{h}(b_\beta, \boldsymbol{\eta}) \right] d\boldsymbol{\eta}, \quad (101)$$

where $\boldsymbol{\eta} > 0$ means that the integral extends to all the Fourier modes which have $\eta_{x_k} > 0$ (half of the Fourier plane). A necessary condition for those class of coefficients being equal to 0 is $\tilde{h}(b_\beta, \boldsymbol{\eta})$ being real. On the other hand, $h(b_\beta, \mathbf{r})$ having reflection symmetry on both axes, is a sufficient condition for that to happen, which is the case for our current model (formula 26).

- The third case is when we are dealing with the position parameters only.

$$\frac{\partial^2 \{-\ln(L_{H_1})\}}{\partial x_k \partial x_j} = 4\pi^2 \int_{\boldsymbol{\eta}} \eta_{x_k} \eta_{x_j} \mathcal{B}(\boldsymbol{\eta})^{-1} \tilde{h}^2(b_\beta, \boldsymbol{\eta}) d\boldsymbol{\eta}. \quad (102)$$

As the example given above, if $h(b_\beta, \mathbf{r})$ has reflection symmetry on both axes, this is a sufficient condition for (where we are ignoring the trivial case $h(b_\beta, \mathbf{r}) = 0$)

$$\frac{\partial^2 \{-\ln(L_{H_1})\}}{\partial x_k \partial x_j} \equiv \begin{cases} = 0 & k \neq j \\ \neq 0 & k = j. \end{cases} \quad (103)$$

to happen. This third set of the Fisher coefficients will be named as $\equiv P$.

Ordering the parameters by collecting the position parameters x_k in the upper left corner, the Fisher matrix will then exhibit a structure of the form displayed ($\boxtimes \equiv$ in general a non zero value).

$$F_{ij} = \begin{bmatrix} \boxtimes & 0 & & & \\ & P & & & 0 \\ 0 & & \boxtimes & & \\ & & & \begin{bmatrix} \boxtimes & \dots & \boxtimes \\ & \ddots & \\ \vdots & & C & \vdots \\ \boxtimes & \dots & & \boxtimes \end{bmatrix} & \\ & 0 & & & \end{bmatrix} \quad (104)$$

One may take advantage of the depicted symmetry of the Fisher matrix when trying to compute its inverse.

$$F_{ij}^{-1} = \begin{bmatrix} F_{xx}^{-1} & 0 & 0 \\ 0 & F_{yy}^{-1} & 0 \\ 0 & 0 & [C^{-1}] \end{bmatrix} \quad (105)$$

As an illustrative example and again, with the help of formula (34), one may explicitly write the error bars on the position parameters

$$\Delta X = \Delta Y = \frac{1}{2\pi^{3/2} \sqrt{\int_0^\infty \eta^3 \mathcal{B}(\eta)^{-1} h^2(b_\beta, \eta) d\eta}}, \quad (106)$$

where we have employed the circular symmetry of proposed template (26).

7.4 Extending to real scenarios

So far we have only considered an extremely simplified model. Our purpose was two fold:

- Make the problem practical in terms of required resources.
- Avoid obscuring and cluttering the fundamentals with complex "decoration" details.

However, we should always keep in mind that the experimental setup we wish to address in the future, the Planck Surveyor satellite, and the object of its study, the full sky, adds other significant challenges as well. Three of them are worth of further discussion:

The in-homogeneity of the statistical properties of the background: When introducing the galactic emission which is mainly limited to a narrow band around the galactic plane, a severe breakdown on the condition of statistical homogeneity occurs. Moreover the instrumental pixel noise is never completely homogeneous. As result of the scan strategy there are zones of the sphere which will be sampled a greater number of times than others, increasing the integration time and leading to lower noise levels than those of the areas where beam spends less time. Using small patches ($\sim 128/256 \times 128/256$ pixels) and re-computing the background for each one of them will decrease this problem to a level which can be safely ignored. Though, it must be noted this is not a limitation of the Bayesian framework which, on theoretical grounds, is prepared for dealing with in-homogeneous backgrounds equally well. We only take advantage of the circulant properties of the covariance matrix, transforming to Fourier space as an implementation technique which allows us a much simpler and fast solution to our problem rendering it practical and efficient. However, the same set of assumptions is required in the derivation of the optimal linear filters together with an implicit assumption of Gaussianity (HM03).

The non-Gaussianity of the galactic diffuse foregrounds: The galactic diffuse fields, for instances dust, which can show considerable values of non-Gaussianity, as result of the Central Limit Theorem, only introduce extremely small distortions on the Gaussian distribution of the Fourier modes. The non-Gaussian part of the field brings in no more than small correlations between the different Fourier modes which can be securely ignored when compared for instances with the potentially much more serious non-homogeneity of the background (HM03) (Rocha et al.(2005)).

The "confusion" noise as result of the contribution to the background of the faint unresolved point sources: This is the dominant component of the background of the Planck high frequency (545 and 857 GHz) maps on high galactic latitudes. This component may be modelled as a Poissonian sampling process with a constant rate source. Any realistic background model should take this into account.

8 RESULTS

We tested the presented ideas performing an extensive set of simulations using two different toy models and different detection scenarios (the latter solely for the white noise only model (see subsection 8.2)). In both cases we used a Gaussian source template (6). The parameters for each source were drawn from uniform distributions with bounds given in table 4. The patches are square grids of 200×200 pixels with a border of $2 \times$ the largest source radius, in pixels. The sources were spread uniformly throughout the patch. Care was taken to avoid source clustering and the presence of sources in the patch's borders. We imposed a safe distance between sources by defining a "bounding" square, centred on the source with side $= \gamma \times 2 \times R$ pixels wide, where $R \equiv$ "Radius of object" (see table 4 for the values of γ). We forbade the overlapping of these "bounding" squares.

Type of simulation	LISNR-16-fix	LISNR-16-var	HMCL
$\langle ISNR \rangle$	3.84	3.62	4.68
ϕ	0.75	0.66	0.66
N_{simul}	5000	1000	10000
% detections	67.41 %	56.41 %	82.95 %
% spurious	9.60 %	8.62 %	8.19 %
Total error	42.19 %	52.20 %	25.15 %

Table 3. Summary of the results from the simulations with a white noise background. N_{simul} is the number of patches which have been simulated, ϕ is the tuning parameter

8.1 Performance of the algorithms

We evaluated the performance of the different algorithms by computing the “*Total Error*” defined as:

$$‘Total Error’ = \text{Percentage of ‘undetected sources’} + \text{Percentage of ‘spurious detections’}$$

where:

- A ‘undetected source’ happens when: a simulated object is not recognized by the algorithm as a source.
- A ‘spurious detection’ happens when:
 - An object is detected
 - There is no simulated object such that the entire set of its parameters belongs to the volume defined by the estimated parameters of the detected object $\pm 3\sigma_p$.
- The ‘*Percentage*’ is computed dividing the quantities by the total number of sources in the patch.

The algorithm that minimizes the ‘*Total error*’ is the one that performs better.

8.2 White noise background

In Table 3 we show the performance of PowellSnakes, where, ϕ is the tuning parameter. The sources are imbedded into a background which is a stationary Gaussian random field with a coherence length much smaller than one pixel, usually known as Gaussian white noise process.

Two scenarios were tested:

- HMCL (Hobson-McLachlan). With this example we tried to mimic, as exactly as possible, the HM03 “Toy problem 4.3”
- LISNR (Low Integrated Signal to Noise Ratio). This example was inspired in another from HM03, where a SZ example was given, but we have only retained the range of the *ISNR* involved. We did not try to simulate the SZ case at all as we have always worked with the Gaussian source template.

Two situations were investigated:

- With a fixed number of sources per patch: “**fix**”
- With a variable number of sources per patch: “**variable**”. In this case the number of sources for each patch was drawn from an uniform distribution whose minimum limit was zero and the maximum was the value displayed in table 4 (N_{objs})

Table 4 has a summary of the parameters’ values, background and foregrounds (sources), for each of the simulated examples. Figure 1 represents a typical likelihood manifold for this case (A and R dimensions suppressed).

Each run (map complete cleanup) takes less than a second (~ 2 each second) on a regular PC (Pentium IV, 2.39 GHz). A considerable effort to optimize the code was made. An intelligent management of previously computed values was essential in achieving such an high performance. Provisions for parallelizing the code were made but not yet employed.

8.2.1 Discussion

One of our major goals when we started this project was making Bayesian methods run fast enough in order to allow us to collect significant statistics figures by performing massive simulations in a reasonable amount of time. We have fulfilled our goal, as you may verify in table 3, several thousands of simulations were performed in each proposed scenario making our figures statistically sounding. Our results are consistent with the theoretical framework presented here. Examples with lower *ISNR* show higher levels of “*Total error*” as expect and significantly higher than the lower bound (see subsection 6.3). Cases with a fixed number of sources per patch always show a lower error. This should not be a surprise as in our priors we have assumed a constant expected number of sources per patch and the lower the dispersion in the number of sources the better the prior fitted the data, hence the higher performance. Worth of a reference is the “*Tuning parameter*” (subsection 6.4). In table 3 we are only showing results achieved using the “*Tuning parameter*” values which performed better. All the

Detection models			
	LISNR-16-var	LISNR-16-fix	Hobson-McLachlan ⁶
R_{max} (pixels)	5.4	5.4	10
R_{min} (pixels)	3.21	3.21	5
A_{max}	0.76	0.8	1
A_{min}	0.2	0.22	0.5
σ	1	1	2
N_{objs}	16 (variable)	16 (fixed)	8 (variable)
γ	1.5	1.5	2.5

Table 4. Detection models parameters, where the R_{max} and R_{min} are the maximum and minimum radius respectively; and A_{max} and A_{min} are the maximum and minimum amplitude respectively; The radii units are pixels and the source amplitudes units are given in the same unit as σ

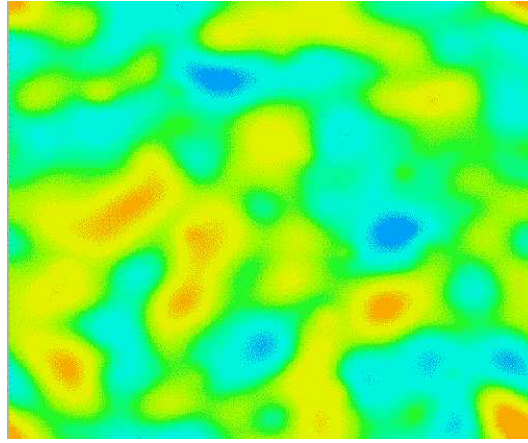


Figure 11. “Coloured” background - astronomical component; Power spectrum was modelled using formula (86); $s_0 = 11.57$ pixels

presented examples are low ISNR cases, thus we have only employed values lower than 1 (see table 3). These “best choice” values are consistent with our predictions.

8.3 “Coloured” background

Our arrangement for the “coloured” background simulations followed the same guidelines as those for the white noise only case. A greater care concerning the realism of the setup was taken into consideration. We used a pixel size of $1' \times 1'$, and small patches of 200×200 pixels ($\sim 3.33^\circ \times 3.33^\circ$). The background consisted of two uncorrelated components:

- A stationary Gaussian random field with a Gaussian power spectrum. The power spectrum, already smoothed by the antenna, was modelled by formula (86) with a coherence length scale parameter of $s_0 = 11.57$ pixels (see figure 11)
- Pixel noise with a “white” profile with constant level across the map
- The CMB RMS level used was $10\mu K$ and the pixel noise $20\mu K$ which together implied $w_I = 4/5$ and $w_B = 1/5$ (see expression 87, figure 12)

The simulated sources had amplitudes ranging from 0.9 to 1.0, in normalized units (normalized by the total rms noise level). The radii of the sources were inside the interval $[3.2, 5.0]$ pixels. The whole setup results in an average $\langle ISNR \rangle$ of $\simeq 3.84$. In each simulated map we put 16 objects with their parameters drawn from uniform distributions (whose limits are given above). Following the same prescription as for the white noise only case, we prevented the sources from forming clusters and being placed on the patch borders (see figure 13). Table ?? contains the results obtained. We have performed a total of 5000 simulations.

8.3.1 Discussion

When we introduced correlation in the background, the likelihood manifold has now considerably changed from the white noise only example. When keeping $\{A, R\}$ constant, looking at the position subspace one may see that the likelihood maxima are now extremely narrow and

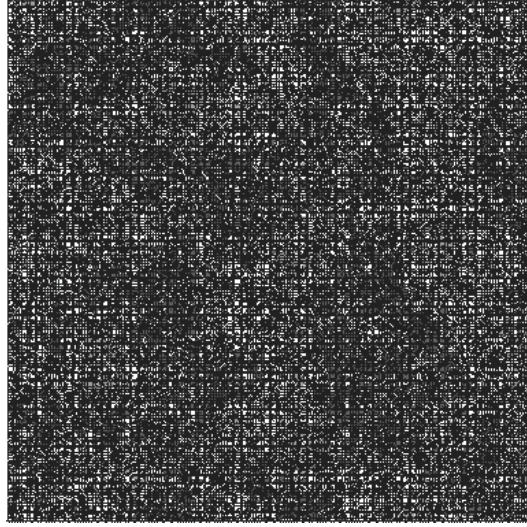


Figure 12. “Coloured” background - complete; astronomical component $10 \mu K$, pixel noise $20 \mu K \Rightarrow w_I = 4/5, w_B = 1/5$

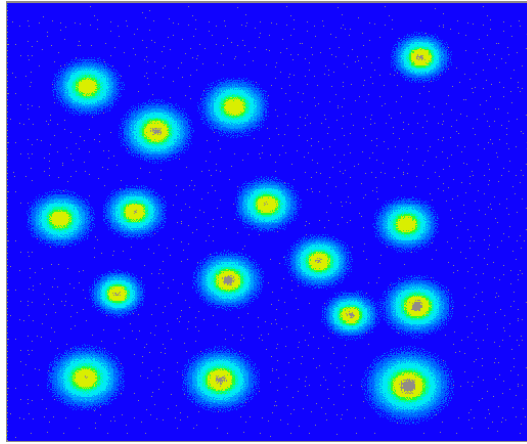


Figure 13. The simulated sources. The shape of the sources was assumed Gaussian (6); all the parameters, $\{A, R, X, Y\}$, were drawn for uniform distributions (check the text for details)

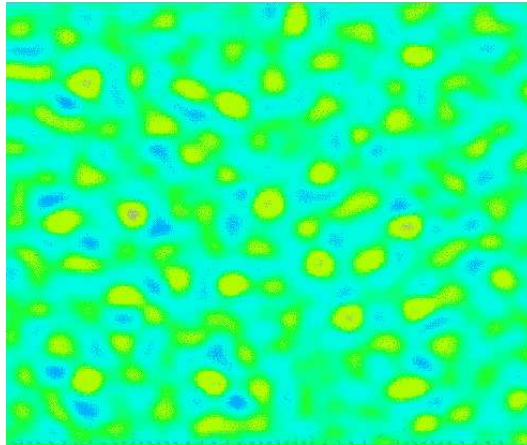


Figure 14. Likelihood ratio $\frac{L_{H1}}{L_{H0}}$ (A and R dimensions suppressed) top view; The sources are the same as in figure 13; the background parameters are those from the simulations: $s_0 = 11.57$ pixels, $w_I = 4/5$, $w_B = 1/5$, $\langle NISNR \rangle \sim 3.84$

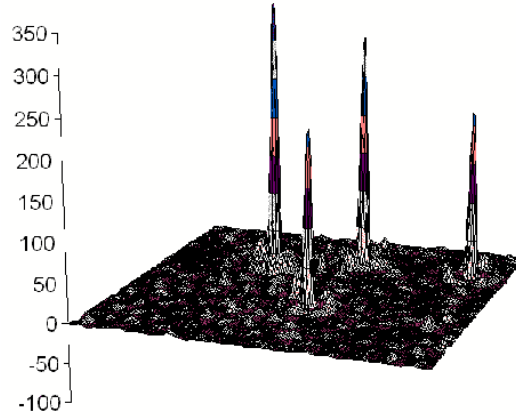


Figure 15. “Coloured background” - The likelihood ratio $\frac{L_{H1}}{L_{H0}}$ manifold position subspace. Example where the instrumental noise is very low; background parameters: $s_0 = 11.57$ pixels, $w_I = 0.01$, $w_B = 0.99$, $NISNR \sim 23.31$

high and they have become surrounded by subsidiary peaks of considerable height (see figure 2). This scenario is even more acute when the proportion of instrumental noise becomes lower in relation to the “coloured” component. In figure (15) we depict an exaggerated case (for clarity) where the instrumental noise is only 10% of the astronomical component. Nevertheless we should stress that current detection instrumentations are already getting very close to these values. The Planck mission LFI instrument (Planck “Blue book” (2005)) has a previewed average ratio of $\sim 20\%$ between the astronomical components and the detectors instrumental noise. Though a more balanced proportion should be expected on the majority of existing instrumental setups. When dealing with such scenarios the search for the likelihood maxima becomes harder. The likelihood peaks being thin and completely surrounded by a very high number of other fake peaks are extremely difficult to find. One now needs to unfold a much higher number of “snakes”, typically four times the number of the white noise only case. The use of the “jump procedure” is now mandatory in order to pass through the large accumulation of fake peaks which encircles the real ones. Since the likelihood peaks are now extremely thin, the error bars on the source localization are now extremely small, usually much less than a pixel. This is the equivalent of assuming that the estimated values for the position parameters perfectly match the true ones. When we start employing the Bayesian procedure for validation of the detection we should replace the flat position prior (see 52) by a Dirac delta centred on the values previously estimated (see 107) for the position of the source:

$$\pi_0(\mathbf{a}) = \frac{\delta(x - x_0)\delta(y - y_0)}{\Delta A \Delta R} \quad (107)$$

As a final note on the speed of the code execution we must note the substantial decrease of performance when dealing with the “coloured background”. The average time for completing a full patch is now about 8 seconds which is about an order of magnitude more than in the previous case. Despite this significant reduction, our initial purpose (performing massive amount of simulations) has not been compromised as one can still undertake thousands of simulations in a reasonable amount of time.

A summary of the sequence of steps performed by PowellSnakes, Version I (the version here exposed), is given in Appendix A.

9 CONCLUSIONS

The detection and characterisation of discrete objects is a common problem in many areas of astrophysics and cosmology. When performing this task most methods assume a smoothly varying background with a characteristic scale or coherence length much larger than the scale of the discrete objects under scrutiny. However high-resolution observations of the cosmic microwave background, CMB, pose a challenge to such assumptions. CMB radiation has a coherence scale length of the order of ~ 10 arcmin, close to that of the objects of interest such as extragalactic ‘point’ sources or the Sunyaev-Zel’dovich effect in galaxy clusters. In addition the instrumental noise levels can be greater than the amplitude of the discrete objects. The common approach for dealing with such difficulties is to apply a matched filter to the initial map and instead analyse the filtered map. These approaches have reported good performances. However the filtering process is only optimal among the limited class of linear filters and is dissociated from the subsequent object detection step of selection performed on the filtered maps. Hobson & McLachlan ((2003); HM03) were the first to introduce a Bayesian approach to the detection and characterization of discrete objects in a diffuse background. Its implementation was performed using Monte-Carlo Markov chain (MCMC) sampling from the posterior which was very computationally expensive. This prompted us to explore a new, fast method for performing Bayesian object detection in which sampling is replaced by multiple local maximisation of the posterior, and the evaluation of errors and Bayesian evidence values is performed by making a Gaussian approximation to the posterior at each peak. In this paper we presented such a novel approach, PowellSnakes. We start by pre-filtering the map and explicitly show that for a given value of R (radius of the source), peaks in the filtered field correspond to peaks in the log-likelihood considered as function of the position of the sources. We replace the standard form of the likelihood for the object parameters by an alternative exact form that is much quicker to evaluate. Next we launch N downhill minimizations in the 2-dimensional (X, Y) using the Powell algorithm. In order to avoid local maxima in the likelihood surface we devised the so called ‘jump procedure’, a

new concept borrowed from the realm of quantum mechanics. Next we perform multiple Powell minimizations in the full 4-dimensional space (X, Y, A, R) , starting each minimization from the positions found in the previous step. To evaluate the performance of an algorithm we introduced here for the first time the concept of ‘symmetric loss’ within the context of the proposed subjects. The Peak-based prior, P , was proposed and an upper bound on the quality of a detection presented. We showed that the quantity, ‘Normalised Integrated Signal to Noise Ratio’, NISNR, plays an important role when establishing a boundary on the source fluxes that can be reliably detected. We included a discussion on the truncation of the posterior by the prior and proposed a novel solution by introducing the ‘tuning parameter’, ϕ . We studied in detail the consequences of adding correlation into the diffuse background which makes the substructure where the sources are imbedded. As the Linear filter works as a whitening equalizer, the coloured noise case reduces to the white noise as long as we use the ‘whitened’ map and sources. Using simulations of several typical scenarios we showed that our approach performs very well on both white and coloured background. We found that our results are consistent with our theoretical framework. For the white noise background, cases with a lower, ‘Integrated Signal to Noise Ratio’, ISNR, exhibit higher levels of ‘Total Error’ as expected. Cases with a fixed number of sources per patch show a lower error. The number of spurious detections increases with this prior however the number of detected sources increases by a larger amount resulting in a lower ‘Total error’. When we introduce colour in the background the likelihood changed considerably: the maxima in the position subspace are now extremely narrow and high surrounded by subsidiary peaks of considerable height. The search for the likelihood maxima becomes harder: one has to resort to a larger number of ‘snakes’. The usage of the jump procedure becomes mandatory. In this case the errors in the source localization are very small, usually much less than a pixel. This means that we can replace the flat position prior by a Dirac delta function centred on the values previously estimated.

Furthermore this approach yields a speed-up over sampling-based methods of many 100s, making the computational complexity of the approach comparable to that of linear filtering methods. Here we presented PowellSnakes, in its first incarnation: ‘PowellSnakes I’. An account of an updated implementation, ‘PowellSnakes II’ will be published shortly. The application of the method to realistic simulated Planck observations will be presented in a forthcoming publication.

ACKNOWLEDGEMENTS

PC thanks the Cavendish Astrophysics Group of the University of Cambridge for support and hospitality during the progression of this work. GR acknowledges support from the US Planck Project, which is funded by the NASA Science Mission Directorate. GR would like to acknowledge useful discussions with Krzysztof Górsky and Charles Lawrence.

10 APPENDIX A

PowellSnakes is a new fast Bayesian approach for the detection of discrete objects immersed in a diffuse background. This new method speeds up traditional Bayesian techniques by:

- replacing the standard form of the likelihood for the parameters characterizing the discrete objects by an alternative exact form that is much quicker to evaluate;
- using a simultaneous multiple minimization code based on Powell’s direction set algorithm to locate rapidly the local maxima in the posterior; and
- deciding whether each located posterior peak corresponds to a real object by performing a Bayesian model selection using an approximate evidence value based on a local Gaussian approximation to the peak. The construction of this Gaussian approximation also provides the covariance matrix of the uncertainties in the derived parameter values for the object in question.

This new approach provides a speed up in performance by a factor of ‘hundreds’ as compared to existing Bayesian source extraction methods that use MCMC to explore the parameter space, such as that presented by Hobson & McLachlan (2003). The method can be implemented in either real and Fourier space. In the case of objects embedded in a homogeneous random field, working in Fourier space provides a further speed up that takes advantage of the fact that the correlation matrix of the background is circulant. Its performance is found to be comparable if not better to that of frequentist techniques such as applying optimal and wavelet filters. Furthermore PowellSnakes has the advantage of consistently defining the threshold for acceptance/rejection based on priors, the same cannot be said of the frequentist methods.

Let’s start by recapitulating that the *Powell’s* method is the prototype of “direction-set methods” for maximization or minimization of functions. These are the methods to apply when you cannot easily compute derivatives. However this method requires a one-dimensional minimization sub-algorithm, eg. *Brent’s* method. The *Brent’s* method is an interpolation scheme whereby you alternate between a parabolic step and golden sections. It is used in one-dimensional minimization problems without calculation of the derivative. (for more details see Numerical Recipes (1992), pg 389, 395, 406).

The steps followed in PowellSnakes, Version I, are:

- (i) Compute the classical linear matched filter $\hat{A} = \frac{t(\mathbf{x})^T N^{-1} \mathbf{d}}{t(\mathbf{x})^T N^{-1} t(\mathbf{x})}$ for the whole patch in the Fourier domain - the denominator is constant and needs to be evaluated only once which can be done in Fourier space as well
- (ii) Compute all the constants
- (iii) Fourier transform the map - a 4Mpix map takes at most 3 secs (FFTW)
- (iv) Filter the map ($\simeq 2$ secs)
- (v) transform back to the map space using the inverse fourier transform ($\simeq 3$ secs)

- (vi) Evaluate the $\ln(\frac{P(H_1|\mathbf{d})}{P(H_0|\mathbf{d})})$ map ($\simeq 2$ secs)
- (vii) Find the peaks of $\ln(\frac{P(H_1|\mathbf{d})}{P(H_0|\mathbf{d})})$ field - check if they are higher than a certain acceptance level: find the peaks using a Powell minimizer with jumps. The evaluation of the objective function is fast - use a kind of evaluation cache (pick a value from a precalculated array). However we still need to find the optimal radius, positions and amplitude in real space. When using the matched filter an average value for the radius of the source has been used and the values of the peaks are only approximate. Now we need to find the optimal parameters.
- (viii) Error bars are calculated from approximate versions of the analytic formulae (while MCMC provides these errors automatically)
- (ix) The stop criteria - no stop criteria - however there is a way of ensuring that no peak is left behind - we proceed as follows: first of all we can have an idea of the area (not the volume) under the likelihood peaks. Let's consider now a imaginary pixel of the size of this area. Split the map into this imaginary pixels and start a PowellSnakes search in the center of each one of them. It is unlikely to miss any of the significant likelihood maxima (specially if we use the jump technique). One might think this process to be time consuming but in fact this is not the case since the objective function is pre-calculated.

10.1 Characteristics

Let's now characterize PowellSnakes in terms of its speed and sensitivity.

10.1.1 Speed

The PowellSnakes algorithm is a two step process, in the first part of the algorithm we proceed as follows:

- (i) First the map is filtered with a linear matched filter given by:

$$\hat{A}(\mathbf{X}, R) = \frac{\mathbf{t}^T(\mathbf{X}, R)\mathbf{N}^{-1}\mathbf{d}}{\mathbf{t}^T(\mathbf{X}, R)\mathbf{N}^{-1}\mathbf{t}(\mathbf{X}, R)}. \quad (108)$$

The evaluation is done in Fourier space using the average radius of the possible range of radiae. The map is Fourier transformed again back to real space. If the radiae of the objects were all equal, and if we knew their true value, the resulted filtered field would be proportional to the likelihood field with the absolute maxima on top of the objects. When the available range of radiae is small, filtering with the 'average' matched filter will not produce absolute maxima but in the great majority of the situations a local peak is created over the objects.

- (ii) Next we search this map with the Powell minimizer in two dimensions (the position) with the jumps switched on. This is an extremely fast operation because the objective function is already pre-calculated, and we just need to pick up a value from an array, or better, compute a bilinear interpolation.

In the second part of the algorithm:

- (i) Now we already have a good idea of where to look further: Around the maxima of the likelihood restricted to the subspace of the position of the objects. In step 1) we produced a 'whitened' version of the map, and we 'whitened' the test objects, at the same time. At this point we switch off the jumps, and launch 'PowellSnakes' in all four dimensions (position, radius and amplitude), taking as starting points the previously found maxima restricted to the position subspace. Use the complete expression for the likelihood (no analytical shortcuts).
- (ii) Once Powell minimizer finishes, pick up the optimal values for the parameters and test them for acceptance/rejection. If and only if the maximum is accepted as 'good', subtract the object from the map. (This is different from the our old solution: sometimes, when we subtracted a rejected peak in the neighbourhood of a true but still undetected object, we ended up damaging the good peak. Unfortunately most of the times that is enough to prevent the real object to be detected. Right now we only subtract accepted peaks to avoid multiple detections of the same object. We think we should try to avoid the subtraction at all, since there is always the risk that a spurious detection and consequent subtraction will damage a good peak.)

10.1.2 Sensitivity

Now, let's consider the Sensitivity:

- (i) Analytical solutions are in general great, however when noise is taken into consideration the functions become hard to handle. Furthermore most of the theorems don't apply in this case. Therefore one should try to avoid analytical solutions.
- (ii) Sometimes the estimated values of the parameters fall off the allowed range used to define the Bayesian criterion for acceptance. Therefore even if the error bars allow them, we don't accept them instead send them for testing (accept/reject), or reject them immediately. Rather you should attach them a very low probability and let the computation continue.
- (iii) PowellSnakes depends on several parameters in order to optimise its performance / sensitivity: ie, the number of launched "snakes", average number of sources per patch, etc. Since the detection conditions (background, noise, signal-to-noise ratio, etc.) can exhibit some drastically changes across the sphere, these parameters should be occasionally re-adjusted. In its current implementation PowellSnakes allows the definition of sphere zones. Within each zone a different set of these parameters may be loaded from the parameter file in order to tune its sensitivity / performance.
- (iv) Do not expect miracles: If the NISNR is low you cannot detect reliably.

10.1.3 Other technical details

(i) The structure of the program was designed to be ready to take full advantage of a multiprocessor system. The performance will scale linearly with the number of processors.

(ii) Implemented in ISO C++.

(iii) Completely built on templates. This feature allows you to consider the precision you want.

(iv) Designed as a ‘framework’, ie you do not need to understand how PowellSnakes works to use it. Inherit from its classes, customize it, and PowellSnakes will call your code when needed.

REFERENCES

- Teukolsky, W.H., Vetterling, S.A. , 1992, Flannery, B.P., ‘Numerical Recipes in C’, CUP, second edition, Press, W.H.
Hobson, M.P., McLachlan, C. , 2003, MNRAS, **338**, 765-784
Jaynes, E. T. , 2003, ‘Probability Theory : The Logic of Science’, CUP, edited by G. Larry Brethorst
Van Trees, H. L. , 2001, ‘Detection, Estimation, and Modulation Theory, Part I’, Wiley-Interscience
López-Caniego et al. , 2005, EURASIP Journal on Applied Signal Processing, Vol 2005:15, 2426.
Cohen-Tannoudji, C. et al. , 1973, ‘Mecanique Quantique’, Hermann
Goodman, J. W. , 2000, ‘Statistical Optics’, Wiley Classics
Hobson, M.P., Lahav, Bridle, S. , 2002, MNRAS **335**, 377
Sanz, J. L. et al., 2001, ApJ **552**, 484
E. Bertin, S. Arnouts, 1996, ‘SExtractor: software for source extraction’ , Astronomy and Astrophysics **117**.
Haehnelt & Tegmark, 1996, MNRAS **279**, 545-556
Herranz et al., 2002, Astrophys.J. **580**, 610-625
Herranz et al., 2002, MNRAS **336** , 1057
Herranz et al., 2005, MNRAS **356** , 944-954
Rocha et al., 2005, MNRAS **357** 1-11
Feroz & Hobson, 2008, MNRAS **384** , 449F
Savage & Oliver, 2007, ApJ **661**, 1339S
ESA, ‘Planck Bluebook’ , 2005, http://www.rssd.esa.int/SA/PLANCK/docs/Bluebook-ESA-SCI%282005%291_V2.pdf